DOCUMENT RESUME

ED 242 290 IR 010 999

AUTHOR	Lancaster, F. W.
TITLE	Guidelines for the Evaluation of Training Courses, Workshops and Seminars. Second Edition.
INSTITUTION	United Nations Educational, Scientific and Cultural Organization, Paris (France). Bureau of Studies and Programming.
REPORT NO	BEP-83/WS/3
PUB DATE	Sep 83
NOTE	139p.; For the 1975 edition, see ED 216 689.
PUB TYPE	Guides - Non-Classroom Use (055) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Adult Education; Cost Effectiveness; Course Evaluation; *Data Collection; *Evaluation Methods; Formative Evaluation; Guidelines; *Professional Continuing Education; *Program Effectiveness; *Program Evaluation; Questionnaires; Reliability; Seminars; Summative Evaluation; Workshops
IDENTIFIERS	*Evaluation Reports; UNESCO

ABSTRACT

This extensive revision of a 1975 publication, "Guidelines for the Evaluation of Training Courses, Workshops, and Seminars in Scientific and Technical Information and Documentation," suggests possible approaches to the evaluation of the type of training sessions that might be organized or promoted by Unesco. Guidelines relate to what data to collect and how, in general, they may be collected; the focus is on adult education and, in particular, on the continuing professional education of adults and on evaluation of program effectiveness, though some attention is given to cost effectiveness or cost benefit aspects. The introduction provides definitions and discusses levels of evaluation, approaches to evaluation, and evaluation data. Additional chapters examine reaction (student reaction, independent observers, and reaction of instructors), and evaluation of learning, behavioral change, program results, specific educational activities, and cost effectiveness. In addition, suggestions are offered for designing evaluation instruments and reporting, reliability and related concerns are addressed, and conclusions provide additional comments on evaluation objectives and methods. A 76-item reference list is included and examples of data collection instruments and evaluation reports are appended. (LMM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Evaluation Training Series
Formation en matière d'évaluation

BEP/83/III
EVALUATION
Paris, September 1983

ED242290

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

GUIDELINES FOR THE EVALUATION OF TRAINING COURSES, WORKSHOPS AND SEMINARS

by

Prof. F.W. Lancaster
Graduate School of Library and Information Science
University of Illinois

Second Edition

BEP-83/WS/3

Bureau of Studies and Programming
Bureau d'études et de programmation

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Suzanne Adlung

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR010666

Reproduced by permission of

UNESCO

CONTENTS

List of exhibits ii

Preface 1

0. Introduction 2
0.1 Definitions 2
0.2 Levels of evaluation 2
0.3 Approaches to evaluation 5
0.4 Evaluation data 11

1. Evaluation of reaction 15
1.1 Student reaction 15
1.2 Independent observers 29
1.3 Reaction of instructors 30

2. Evaluation of learning 33

3. Evaluation of behavioural change 38

4. Evaluation of programme results 44

5. Design of evaluation instruments 49

6. Evaluating specific educational activities 65

7. Evaluating cost-effectiveness 73

8. Reporting 77

9. Reliability and related concerns 80

10. Conclusion 82

References cited 85

Appendices
1. Examples of questionnaires used in ongoing evaluation 91
2. Example of a checklist for use by an independent observer 96
3. Interview guide used in follow-up telephone interviews 97
4. Follow-up questionnaire 101
5. Examples of questionnaires used in terminal evaluation 107
6. Follow-up questionnaire developed for UNISIST course 113
7. Examples of reports of evaluations 115

List of Exhibits

1. Levels of evaluation applicable to a course 3

2. Classification of approaches to evaluation in education 12

3. Major categories of data in the evaluation of education 13

4. Types of data to be collected in the ongoing formative evaluation of a course 20

5. Daily log 21

6. Single session check sheet 22

7. Some important questions to be answered in the terminal evaluation of a course 23

8. Evaluation instrument proposed by Bryant 25

9. Student evaluation of the extent to which course objectives were met 26

10. Excerpt from the diary of an evaluation observer 29

11. Instructor evaluation forms 32

12. Example of statements for testing factual knowledge 35

13. Follow-up questionnaire 41

14. Example of interview with participants six months after conclusion of a course 43

15. Instructions to participants in a course on position classification 47

16. Sample page from student self-assessment 48

17. Closed questions from a course evaluation questionnaire 53

18. Open-ended questions from a course evaluation questionnaire 54

19. Sample of data from a course evaluation questionnaire 55

20. Typical numerical scale used in evaluation instrument 56

21. Use of the Smiley chart 57

22. Use of the semantic differential 59

23. The Kropp-Verner scale 60

24. Some of the behavioural objectives developed for a course in information storage and retrieval 68

25. Pre-course background questionnaire 70

26. Effectiveness/feasibility considerations relating to four instructional alternatives 74

27. Cost analysis of the four alternatives identified in Exhibit 26 76

28. Practical suggestions on reporting evaluation results 78

29. Concise format for presenting evaluation data designed to answer a single question 79

PREFACE

The first edition of this publication was issued by Unesco in 1975 under the title Guidelines for the Evaluation of Training Courses, Workshops, and Seminars in Scientific and Technical Information and Documentation (SC.75/WS/44). Because it was found useful beyond the subject area for which it was originally intended, the author was asked to produce an updated edition. The present version is completely revised and the former subject restriction has been lifted. The justification for this is that the specific subject content dealt with in a training programme is not in any way the major factor influencing the type of approach appropriate to the evaluation of that programme.

INTRODUCTION

The literature relating to evaluation in education is vast and continuing to proliferate rapidly. Indeed, Miranchuk[45] has referred to evaluation as the "newest growth industry in education". There is also a growing literature on the evaluation of evaluation methods in education. Stufflebeam[67] has referred to this as metaevaluation. Much of the literature on evaluation in education is philosophical or theoretical, rather than practical; it relates more to "education" in a broad, abstract sense, or to very large educational programmes, than to the evaluation of specific educational offerings.

In these guidelines no attempt has been made to comprehensively survey the evaluation of education as reflected in this literature. The purpose is a more practical one: to suggest possible approaches to the evaluation of training courses, workshops and seminars of the type that might be organized or promoted by Unesco. The focus is adult education and, in particular, the continuing professional education of adults.

While the guidelines give practical advice on evaluation procedures, they are not intended as a book of recipes. Each evaluation situation differs somewhat from all others. An evaluation approach that works in one setting will not necessarily transfer intact to another. The guidelines are intended to aid in the design and execution of evaluation procedures appropriate to different situations, needs and constraints.

0.1. Definitions

Good[20] has put forward the following definitions:

> Course. Organized subject matter in which instruction is offered within a given period of time, and for which credit toward graduation or certification is usually given.
>
> Seminar. An instructional technique common in but not limited to higher education in which a group of students engaged in research or advanced study meets under the general direction of one or more leaders for a discussion of problems of mutual interest.
>
> Workshop. An instructional method in which persons with common interests and problems meet with appropriate specialists to acquire necessary information and develop solutions through group study; usually residential and of several days duration.

In terms of evaluation, however, these distinctions are not worth maintaining because, although minor variations may exist, the problems of evaluation are mostly common to all three types of educational offering. This report, then, will use the term "course" to refer generically to any type of educational offering.

0.2. Levels of evaluation

Several levels of evaluation that can be applied to an educational programme are suggested in Exhibit 1. The illustration indicates that:

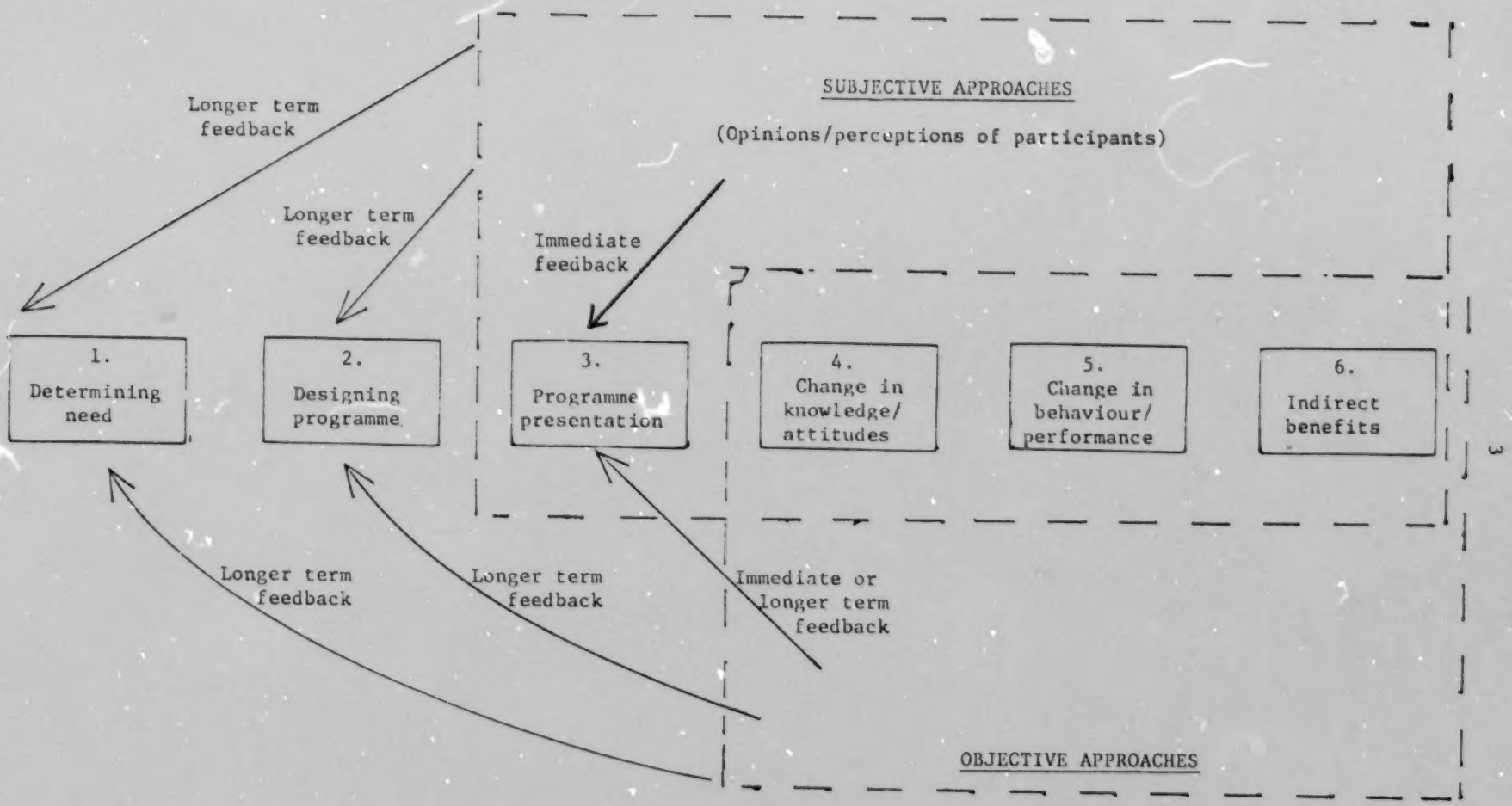


Exhibit 1

Levels and Types of Evaluation

1. The imparting of knowledge of some type to a particular community has been identified as desirable. This recognition is based on some form of needs assessment (a useful, concise summary of needs assessment has been prepared by Campbell[6]).

2. An educational programme (e.g., a half-day workshop) to impart this knowledge is designed and

3. It is implemented.

4. The programme changes the knowledge, skills or attitudes of the participants.

5. Their participation in the programme has some effect in changing their behaviour or performance. For example, as a result of participating in a continuing education programme, a group of physicians may adopt a new approach to the diagnosis or treatment of a particular disease.

6. This change of behaviour produces its own benefits. Hopefully, patients suffering from the disease experience reduced morbidity or mortality as the result of the adoption of improved medical procedures.

From an evaluation standpoint, this situation can be looked at at several different levels. An obvious concern is the quality of the presentation of the programme: were the instructors satisfied with it; were the students? At the next level, an important evaluation question would be "did the course impart the knowledge, skills or attitudes it was designed to impart?" In the hypothetical case alluded to earlier, did the physicians learn enough to be able to apply the procedure correctly in appropriate situations? Beyond this, was the course successful in persuading the physicians to adopt the procedure and, perhaps, to encourage other colleagues to adopt it? The ultimate evaluation criterion, however, is presumably whether or not the health of the community improves as a result of this chain of events.

Moving in the opposite direction, some other evaluation questions emerge: was the programme well designed, was this type of course the best method to use in meeting this particular need, was the need a real one, are there other training needs in the community that are more pressing and that should have been given higher priority?

Evaluation procedures can apply to all the steps involved in a particular educational activity. Moreover, they should be regarded as essential elements in any educational programme. As Hampton[26] points out, "evaluation must occur before, throughout and beyond the entire programme".

Hampton makes a strong case of the importance of evaluation to the field of education:

> Evaluation is one of the most neglected aspects of programming in continuing education today. Yet it has the potential of being one of the most powerful administrative tools available to educators. Most program planners do a respectable job of planning and carrying out their plans to a high degree of proficiency. At this point, however, the program development process tends to break down, for too often planners fail to review their program. Information derived

> from reliable appraisals of educational efforts is essential to the educator in order to rationalize decision-making with respect to improving future programming efforts. Formal and informal judgements made about the program should logically evolve into a plan of operation, which ultimately may affect program process, teaching methods and program accomplishments. (Page 105).

He insists that evaluation should be regarded as an integral part of the educational process rather than just an adjunct. Evaluation is an ongoing activity and one that should be built into an educational programme from the initial planning stages.

Knox[34] identifies four questions that need to be answered in making decisions relating to evaluation: (1) who needs to know how effective the programme is, (2) what do they need to know, (3) who can best provide this information, and (4) how can it be obtained most effectively and efficiently? These matters will be addressed at various points throughout this report.

0.3. Approaches to evaluation

There are many possible approaches to evaluation and many ways in which the subject of evaluation in education can be classified. In general, an activity can be examined in terms of its costs, its quality (effectiveness or performance) and its benefits (impact). Being most tangible, the costs of the activity tend to be easiest to measure. It is usually more difficult to judge the effectiveness of the activity and more difficult still to assess its benefits.

An evaluation to determine how well the activity is satisfying the needs of those for whom it was designed may be termed an evaluation of effectiveness, one that determines whether the needs are being satisfied as efficiently and economically as possible may be termed a cost-effectiveness evaluation, and one that determines whether or not the benefits of the activity exceed its costs may be referred to as a cost-benefit study. Clearly, this sequence is one of increasing complexity.

While this categorization is a valid and useful one, in practice it is frequently difficult to make clear distinctions of this type. In particular, it is often difficult to distinguish cost-effectiveness analysis from cost-benefit analysis and, in fact, these terms tend to be used very loosely. For example, the Joint Committee on Standards for Educational Evaluation[30], in an otherwise useful publication, completely confuses cost-effectiveness and cost-benefit analysis. Costs, performance (effectiveness) and benefits of a particular activity are very closely interrelated and a complete evaluation would take all of these aspects into account (i.e., it would be a cost-performance-benefits study).

The evaluation of effectiveness may itself be subdivided into:

(a) evaluation of how well the activity is carried out;

(b) evaluation to determine how it is carried out and whether or not it can be improved.

The first level can be referred to as macroevaluation. Macroevaluation is concerned with the overall assessment of performance (i.e., how well the objectives of the activity are being met). Macroevaluation in itself tells us how well an activity

is performing but not why the performance is at a particular level, what factors contribute to establishing the performance at this level, what failures occur in the operation, or how the operation might be improved. Macroevaluation alone is a relatively sterile task since, presumably, the principal objective of any evaluation is to produce improvements in the activity being studied. In order to improve an activity one must undertake a more detailed level of analysis to determine how it is carried out, how much failure occurs, why this failure occurs, and what might be done in the future to raise the overall level of performance. This analytical level of evaluation can be referred to as microevaluation. Microevaluation is essentially diagnostic, the objective being to improve the performance of the activity under review.

This categorization of evaluation is as applicable to educational activities as to any other. To return to our hypothetical case, suppose a local medical society sponsors a short seminar to educate all the physicians in a community in the use of some new therapeutic or diagnostic procedure. The effectiveness of the seminar could be evaluated, at the macroevaluation level, by asking the physicians attending how satisfied they were with the methods of presentation. Alternatively, the impact of the seminar could be evaluated more objectively; e.g., by using some kind of test instrument to determine how well informed the physician was in this area before and after attendance. Microevaluation would go beyond this, however, and would seek to discover, in detail, why 38% (say) of those attending were not satisfied with the seminar or why it had no significant impact in increasing the knowledge of a certain proportion of those participating. The objectives of the microevaluation would be to improve this seminar or other seminars of this type in the future. A cost-effectiveness analysis would attempt to determine whether the objectives of the programme (i.e., informing the physician community of a medical innovation) could have been achieved more efficiently or more economically in some other way (e.g., by a videotape made available to physicians in their own offices). The cost-effectiveness of an operation can be improved in one of two ways: (1) by raising the level of effectiveness while holding costs constant, or (2) by maintaining roughly the same level of effectiveness while reducing the cost of achieving this level. The videotape presentation, in this hypothetical case, would be more cost-effective than the seminar if it could be shown to be less expensive to implement but equally effective in transmitting the needed knowledge or, alternatively, if it could be shown to be more effective as a means of transmitting this knowledge while costing no more than the alternative form of presentation. Finally, a cost-benefit analysis would be concerned with programme justification. In this hypothetical case, do the benefits of this particular endeavour outweigh the cost of providing it? One way of approaching this would be to consider whether the medical society should be spending its money on this enterprise or whether it would in fact be more valuable (in terms of benefit to members or to the community at large) to spend the money in some other way. Cost-benefit studies are concerned with value per dollar invested. But the long-term value of many activities (and education is certainly not an exception) is exceedingly difficult to express in absolute terms and it is for this reason that cost-benefit studies are highly complex undertakings if approached in a systematic way. In actual fact, most cost-benefit "analysis" is highly subjective. A decision is made by an individual, or group of individuals, that a particular programme is worth carrying out. That is, they decide that the benefits of the programme outweigh the cost of providing it.

The major emphasis of these guidelines is the evaluation of programme effectiveness, but some reference will occasionally be made to cost-effectiveness or cost-benefit aspects.

A distinction can be made between subjective and objective approaches to evaluation. Subjective evaluation is based on opinions: of students, of teachers, or of an independent observer. Objective evaluation, on the other hand, attempts to move away from opinion, pure and simple, and to arrive at an assessment that is more standardized and perhaps more quantifiable. An obvious example of an objective evaluation is one in which the success of a course is measured by testing the knowledge or abilities of the students before and after their participation. Some standardized test is applied before the course and again when it is completed, the purpose being to measure the change in the students as a result of their participation. Presumably, if the educational experience has been successful, the students will achieve significantly higher "scores" on the second test than they did on the first. A variation on this is the use of some standard test applied to two matched groups of students, each group having been exposed to a different method of presenting the same material, the object here being to compare the success of one approach with that of the other.

Subjective approaches to evaluation are potentially applicable to all the steps identified in Exhibit 1. That is, one can ask the students participating in a course for their reaction to the content and the way it was presented. One could also ask them whether they feel that their knowledge and/or attitudes have changed as a result of participation. Subsequently, questionnaires or interviews could be used to determine, from the course participants themselves, whether they feel that participation in the course has changed their behaviour or performance or has had any indirect benefits. Returning to our hypothetical situation, one could even ask the physicians if they feel that their adoption of a particular procedure has had a significant effect in improving the treatment of some disease among their patients.

There is nothing wrong with getting the opinions of participants in this way. Nevertheless, a more objective approach may be preferable in evaluating the course in terms of knowledge acquired, attitudinal or behavioural change, or indirect benefits. Objective approaches, of course, are more difficult to implement than those that are purely subjective.

Scriven[54] first made an important distinction between formative and summative evaluation. More recently, Studebaker et al.[66] have differentiated among pre-formative, formative and summative evaluations. A formative evaluation of a course is one designed to improve that course. "Dry run" activities, to test out aspects of the course before it is presented to the intended audience, can be considered formative. Formative evaluation can also be applied while the course is in progress, the objective being to improve this particular educational experience before it is completed. To take a simple example, consider a course that lasts for three weeks. It could be evaluated at the end of the first week and again at the end of the second. On the basis of feedback from participants, it may be possible to change the course to make it a more meaningful educational experience in the second and third weeks. That is, one may change the teaching approach or the direction or emphasis of the course content. In the case of a course or seminar of one week duration, one can obtain daily reaction from participants. It is obvious that formative evaluation of this kind is more feasible with courses that are of relatively long duration. Formative evaluation is not necessarily purely subjective. It is perfectly possible to have a formative test of the learning acquired by students. Such a test might, for example, determine whether or not the students can demonstrate the achievement of certain behavioural objectives established for a course. In the case of a course to be repeated several times, an evaluation performed at the end of the course can also be regarded as formative if

it is intended to improve the quality of future offerings in the series. As emphasized by Misanchuk[45], formative evaluation can only apply to a product that is, in some sense, still in a developmental stage.

Summative evaluation, on the other hand, is evaluation of a finished product. Both Misanchuk[45] and Studebaker et al.[66] are careful to point out that summative evaluation is not intended to improve a course but to demonstrate what it can do (i.e., to show its worth). Summative evaluation gathers information for administrators - to aid a decision concerning the future of a course (e.g., adoption, continuation or termination decisions).

The preformative evaluation that Studebaker et al. identify refers to needs assessment, programme planning, and any other evaluative activities that take place before or early in the development of a programme. Preformative evaluation would include:

(1) evaluating the need for various types of activity to allow the establishment of meaningful priorities;

(2) evaluating the goals and objectives of proposed activities to determine if they are, in fact, realistic;

(3) evaluating plans for the achievement of stated goals and objectives.

The distinction between formative and summative evaluation illustrates the fact that evaluation can be conducted from a number of different viewpoints, the most important being those of:

(a) the students

(b) the teachers

(c) those responsible for planning, administering or funding the programme.

The teaching staff itself should be most interested in formative evaluation, the object being to improve the quality of programme content and their own teaching methods. Steele[65] has stated the need for formative evaluation in this way:

> Not only must a constant check be kept on how participants are reacting, but evidence of negative reaction must be fed back into the programme immediately so adjustments can be made. (Page 7)

Course participants should be concerned with formative evaluation; they will want to provide feedback to the teaching staff in order, where necessary, to change the direction or emphasis of the course and thereby make it more responsive to their own requirements. Participants will also want to know how well they are progressing in a course (in terms of their own objectives or the objectives set by the teaching staff) and how they might be able to apply the learning gained. As Knox[34] points out, the participant may also want to discover what he does and does not know relative to the subject of the course - so that he may concentrate on what he does not know. Potential participants in a programme are likely to be interested in evaluations of previous courses organized by the same body and in the same subject area or one closely related. Government agencies and other employers may also find such evaluations useful in deciding whether to send participants to a particular course and, if so, which people to send.

Planners and administrators of educational programmes are likely to want evaluations that are more comprehensive and far-reaching. Although they may want to see evaluation results from individual courses, they will also need to evaluate complete programmes and are likely to be concerned with the total impact of a programme on all participants.

Another way of looking at the evaluation of education is in terms of what is being evaluated. Steele/65 identifies three categories of evaluation when the subject is viewed in this way:

(1) evaluation of the students (e.g., in terms of grades);

(2) evaluation of the teachers (for salary increases, promotion, tenure);

(3) evaluation of educational programmes.

The organizations to whom this report is largely addressed will mainly be interested in the third of these levels: the success or impact of a particular course or of an entire programme of courses. They will not be interested in evaluating the students as such, since most will not be examining bodies or agencies of accreditation, but they might well be interested in the evaluation of students as a means of assessing the success of a programme. For example, one way of measuring the success of a course is in terms of the level of knowledge of the students before and after they participate. Evaluation of the teaching staff will be important to the organizers, not for the reward purposes Steele alludes to, but as part of the overall programme evaluation and in the decision as to whether or not to use the same lecturers again on other assignments. The evaluation of teachers per se will not be dealt with in detail in this report. A very useful discussion of approaches is given by Braskamp et al/73.

Evaluation may be formal or it may be informal. Informal evaluation of an educational offering occurs all the time. A lecturer will solicit feedback from participants at a workshop or seminar, over coffee, at lunch, or on other occasions in which he mixes with students. Even if he does not actively solicit evaluative reaction of this kind, he is likely to receive it gratuitously. Whether or not they are asked to do so, students (consciously or unconsciously) will conduct their own evaluation of any course or workshop they attend. That is, they will make their own judgements about the quality and relevance of the content, the success of the approaches to teaching, and so on. By mixing with students, an independent observer may be able to obtain a fairly clear idea of how the course is going. The teacher, also, is likely to sense how well it is progressing and will be evaluating himself as the course develops. In contrast to this informal or casual evaluation, formal evaluation implies a deliberate effort on the part of the organizers to get some assessment of the success of a course. A formal evaluation will involve some "standardized" approach to data gathering, usually by the use of a questionnaire, interviews with the participants, or some controlled comparison or test.

Another distinction worth making is that between internal evaluation and external evaluation. Internal evaluation is performed by those who conduct the educational programme and it is usually formative. External evaluation is performed by individuals somewhat removed from the programme. External evaluation may be performed for formative or summative reasons.

One of the clearest expositions of approaches to evaluation is that given by Hampton[26], who identifies four possible "steps" proposed earlier by Kirkpatrick[32]. These steps are:

(1) evaluating reaction of participants;

(2) evaluating learning acquired;

(3) evaluating behavioural change;

(4) evaluating programme results.

Clearly, these steps are all illustrated, explicitly or implicitly, in Exhibit 1. Evaluation of reaction (of students, lecturers, observers) is easiest to accomplish. Such evaluation is completely subjective, although the data may be gathered in some systematic way and in a consistent form. These data may also be quantifiable in some sense (e.g., 80% of the participants were satisfied with the approach used in teaching). Evaluation of learning is best achieved through some objective procedure, usually some form of test. More difficult is the evaluation of behavioural change in the participants. This goes beyond learning as such into the application of the learning acquired. It is possible for someone to "learn" some body of knowledge (in the sense that he can pass some test which may be based on memorization) but still not be able to apply it in a practical situation. One approach to the evaluation of behavioural changes is by the measurement of on-the-job performance of an individual before participation in some educational programme and again some time after participation. Preferably this evaluation should be objective. Conceivably, however, it could be subjective, based on the opinions of peers or supervisors. Another possible method of measuring behavioural change is by the use of tests of problem solving or decision making abilities. Programme evaluation differs from the types mentioned earlier in scale rather than in approach or form. The evaluation of an educational programme is of concern to those who plan and administer it. Programme evaluation implies the existence of a set of programme objectives. The evaluation is conducted to determine how well these objectives have been met. Clearly, a complete programme evaluation could involve studies of reaction, learning or behavioural changes, or all three of these, depending on what the objectives of the programme happen to be.

Knox[35] uses a somewhat similar classification of elements that may be evaluated: (1) general impact, (2) participant satisfaction, (3) content mastery, (4) personality (e.g., improved self-image or greater social awareness), and (5) work-related (e.g., on-the-job performance).

Finally, as Hampton[26] again points out, one can categorize evaluation activities by the stage at which they are performed, as follows:

(1) pre-course evaluation,

(2) ongoing evaluation,

(3) terminal evaluation, and

(4) follow-up evaluation.

These evaluation stages are obviously closely related to and identifiable with the various "approaches" mentioned earlier. Ongoing evaluation is likely to be

formative while terminal evaluation may be formative or summative. Follow-up evaluation is likely to be used in the measurement of behavioural changes, while the ongoing evaluation and the terminal evaluation are likely to be based on measurements of reaction and learning. Pre-course evaluation may be conducted for the purpose of judging the knowledge of participants or it may be used as one element in a more complete programme evaluation (i.e., some form of pre-course evaluation, in this case, may be conducted in order to establish meaningful objectives for the programme).

The various ways of classifying approaches to evaluation in education, as identified in this section of the report, are summarized in Exhibit 2.

3.4. Evaluation data

So far, the discussion in this report has largely been restricted to judgemental aspects of evaluation; i.e., evaluation that allows us to make certain judgements about a programme and how far it has succeeded. Stake[60-62], however, has pointed out that descriptive data are also needed in the evaluation of education. The kinds of data to be collected on some educational programmes are viewed as a two-dimensional matrix, as illustrated in Exhibit 3. Antecedents, transactions and outcomes are various characteristics of an educational programme for which data may be gathered. Antecedents are conditions existing before the educational programme, conditions that may affect the outcome, especially the characteristics of the students themselves - their experience, knowledge of the subject matter, aptitudes, and so on. The antecedents are really the inputs to the educational activity, including instructors, materials and planning, as well as the students. Transactions are the activities occurring during the programme, the encounters among the students and teachers, students and students, and students and materials, while outcomes are the results of the transactions, short term and long term, especially evidence of the impact of instruction upon the participants. These three characteristics (other writers, Roid[53] for example, refer to these as inputs, processes and outputs) may be depicted by descriptive data or by judgemental data. According to Stake, intents are conditions that are planned for by the organizers or the instructors, including course content, teaching methods, and intended student behaviour (i.e., what teaching and what learning are intended). Observations are descriptions of what actually did occur, relating to antecedents and outcomes as well as transactions. Judgemental data are also of two types. One can judge various aspects of the programme according to some absolute standards adopted from elsewhere or devised for this specific purpose. Alternatively, one can use relative standards whereby one programme is compared with others.

Stake points out that descriptive evaluation data can be used in two ways: (a) looking for contingencies among antecedents, transactions, and outcomes, and (b) looking for congruency between the intended and the observed. These uses of evaluation data improve our understanding of the effectiveness of the planning and implementation of a programme.

Provus[52] has described an evaluation model very similar to that of Stake. He recognizes four stages of evaluation corresponding to four stages of programme development: definition, installation, process and product. Evaluation moves through these four stages and through three major content categories: inputs, processes and outcomes (which correspond closely to Stake's antecedents, transactions, and outcomes). The definition stage is concerned with programme intents.

Exhibit 2

Classification of Approaches to Evaluation in Education

By purpose	By viewpoint	By object of evaluation	By form	By objectivity	By characteristic measured	By stage
Pre-formative	Student	Student performance	Informal	Subjective	Reaction	Pre-course
Formative	Teacher	Teacher performance	Formal	Objective	Learning	On-going
Summative	Programme planner	Programme performance			Behavioural change	Terminal
					Programme results	Follow-up

Exhibit 3

Major Categories of Data in the Evaluation of Education

as Identified by Stake

	Antecedents	Transactions	Outcomes
Descriptive data a) Intents b) Observations			
Judgemental data a) Absolute standards b) Relative standards			

At the installation stage the evaluator is primarily concerned with observing transactions. At the process stage he responds to observed outcomes and tries to relate the outcomes to the transactions. Finally, at the product stage, he looks for congruencies between the intents and the outcomes.

Elsewhere, Stake[63] has made a distinction between "preordinate evaluation" and "responsive evaluation". Preordinate evaluation is based upon pre-specification in the form of statements of objectives, tests of one kind or another, and perhaps standards of various types. Responsive evaluation, on the other hand, is oriented less to programme objectives than to actual programme activities (the transactions of Exhibit 3). As described by Stake, responsive evaluation is less formal than preordinate evaluation. The evaluator observes the programme in action and interacts with the various groups having an interest in the programme. Responsive evaluation is more descriptive than judgemental.

A categorization similar to that of Exhibit 3 has been made by Knox[35]. The key components in his model are: context (the setting and rationale of the programme), inputs (both people and materials), process and activity (which correspond to the "transactions" of Exhibit 3), outcomes, judgements and application of findings.

Section 0.3 identified several approaches to the classification of evaluation in education. These various classifications, summarized in Exhibit 2, are all closely interrelated (e.g., the purpose or viewpoint will influence the characteristic to be measured and purpose and characteristic to be measured will both determine the stage at which the evaluation is conducted). Exhibit 2 is by no means a complete "classification of classifications". Alternative forms of categorization are possible and many of these alternatives have appeared in the literature. Nevertheless, this table seems to cover the most important ways of looking at evaluation, at least for the purpose of this particular study. The remainder of this report will mainly be concerned with methods of evaluation, relating these methods primarily to the characteristics to be measured.

The major objective is to assist organizations in designing evaluation procedures to (a) improve the quality and relevance of educational programmes, and (b) provide data that will allow those responsible for planning a programme to decide how successful it has been, how far it has contributed to the attainment of stated objectives, and what might be done to improve educational offerings of this type in the future.

1. EVALUATION OF REACTION

1.1. Student reaction

The most important aspect of this form of evaluation is presumably the reaction of the "students" attending the course. Studies of participant reaction will tend to be subjective. That is, methods will be used to determine the opinions of the participants relating to the course in general and, possibly, certain specific features of it. At the most general level, an evaluation of reaction seeks to determine how "happy" the students are with the way a course is progressing or with the way it was conducted. In fact, the type of data collected in this form of evaluation has been referred to as "happiness data" (see, for example, Knowles[33]) or as a "happiness index".

Evaluation of student reaction has definite value. As Hampton[26] points out:

> It is important to know how people feel about the programs they attend, for it is reasonable to expect that participants who enjoy a program are more likely to obtain maximum benefit from it. (Page 107).

Knowles[33] has stated that:

> On the whole, this kind of feedback is most useful in providing a general sense of trends in morale and satisfaction, but it frequently turns up specific and practical suggestions for improvement in the general program or in specific activities; and it may reveal problem points that call for deeper evaluation. (Page 224).

"Reaction data" can be gathered for the purpose of formative evaluation or for summative evaluation. Both formal and informal procedures can be applied. At a more informal level, the teaching staff can ask the students for their unstructured, "off the cuff" impressions of the course. With a relatively small group this could possibly be achieved through an informal discussion with the entire group at the end of each day of meetings. At an even more informal level, it is quite likely that useful reaction data can be obtained through discussions, between students and teaching staff, over coffee, at lunch, and during organized social events. An independent outside evaluator can obtain reaction data in a similar informal way, by observing the progress of a course and by generally "mixing" with students. The outside observer may, in fact, be in a much better position than the teaching staff to gather informal reactions in this way. Participants who may be reluctant to criticize an instructor to his face may be more candid when asked for their opinions by some apparently impartial observer.

A more formal approach to the collection of reaction data will use some type of structured instrument for data gathering. Usually this will be a questionnaire completed by each student, probably anonymously, although interviews may be used in place of the questionnaire. If interviews are used, it is important that they be conducted in some consistent manner, following an interview guide of some kind. Questionnaires, although widely used and accepted as survey instruments in social science research, are criticized by some investigators, usually for two major reasons:

(1) Questions may be misinterpreted by respondents and it is sometimes difficult to know whether or not the respondent has interpreted a particular question in the way the designer of the instrument intended;

(2) There is sometimes some doubt as to whether a respondent has answered truthfully or accurately and there may be no convenient or practical way to check the validity or accuracy of a response.

The first objection is unlikely to be too serious if the number of people attending a course is sufficiently small. It will be possible to have someone in attendance while questionnaires are being completed within the group. This person (one of the lecturers perhaps) will be available to "interpret" the questionnaire to participants and to answer any questions they may have on how to complete it.

The question of veracity or validity of responses is unlikely to apply to a student evaluation of a course. This is a problem that is more likely to apply in a situation in which the respondent is, in some sense, evaluating himself. For example, there may be a tendency for a respondent to overestimate the number of journals read, the number of hours spent reading, or the number of his own publications. This is a matter of prestige and of the desire of the respondent to appear "in the best light". There is no reason to suppose, however, that the participant in a course would want to be untruthful or otherwise inaccurate in the completion of a questionnaire - the nature of the questions asked virtually precludes the possibility of this problem occurring.

Interviews seems to have three major advantages over questionnaires as methods of data gathering:

(1) the presence of the interviewer tends to ensure that all questions are correctly interpreted by the respondent;

(2) it may be possible, by means of "probing" questions, for the interviewer to check on the accuracy of responses;

(3) the interviewer may be able to collect unsolicited observations from the person interviewed; data unanticipated in the interview schedule may thus be collected.

As previously indicated, the first two of these advantages are unlikely to be very significant in the evaluation of courses, and the third benefit does not seem sufficiently important to warrant the use of interviews in place of questionnaires. Interviews are more expensive and time-consuming. They require scheduling of participants, which may not be at all easy to arrange, particularly at the end of a course when most of the attendees are likely to be anxious to leave. Moreover, interviews cannot be completed anonymously, unlike questionnaires, and, for reasons stated earlier, they would require use of an independent interviewer. Members of the teaching staff should not conduct such interviews: they are unlikely to get completely candid responses and they may, perhaps completely unwittingly, influence the responses by the way they pose the questions. For formal measurement of student reaction, then, the printed questionnaire is likely to be the preferred instrument for gathering data.

Nevertheless, where a skilled, independent interviewer is available, there is little doubt that interviews can elicit more detailed and revealing responses. Patton[48] has made a strong case for "qualitative evaluation methods", which seek:

> to understand the perceptions, feelings, and knowledge of people... through in-depth, intensive interviewing. (Page 29).

Patton presents an excellent, detailed description of such qualitative methods. His book is highly recommended as a guide to the conduct of in-depth interviews, as well as to the implementation of observation procedures.

Interviewing of course participants, whenever this is possible, may be particularly appropriate in the international setting: participants with diverse language backgrounds may be better able to articulate their reactions in conversation than in writing.

To allow meaningful evaluation, on the basis of reaction data, it is important that some statement of the objectives of the course be available to participants. One important facet of the evaluation will be the students' assessment of how well the course has met its stated objectives. If no stated objectives exist, or are available to students, the possibility exists that they will evaluate the course in relation to goals that the organizers never intended. While it will also be important to learn of the personal objectives of students, and their judgement on how far these have been met, a wide discrepancy between the objectives of the students and those of the organizers would tend to indicate a failure in the advertising and promotion of the course or, possibly, the selection of students to participate.

But evaluation should not be solely on the basis of stated objectives. A course may have benefits to the participants that were not anticipated by the organizers. In some programmes, for example, the value of the social contact among students may be greater than the value of the formal presentations. It is possible that a course might receive a rather poor score in relation to its stated objectives but still be a valuable educational experience for some other reason. It is important, therefore, that the evaluation of any educational enterprise should be sufficiently open-ended that it will take account of benefits that are apparent to the students although not anticipated by those responsible for planning and organizing the endeavour. Parenthetically, it is worth mentioning that certain educational experiences may also have undesirable consequences that were unanticipated by the organizers or the teachers. For example, it is conceivable that a course on "literacy appreciation" may cause certain students to hate reading, especially if the course requires the student to read a large body of literature in which he may have little or no interest. Harmful effects of this kind are more likely to occur in the basic education of children or young people, and less likely in the continuing professional education of an adult population of the type this report is concerned with. Some useful advice on the identification of unintended effects is given by Wolf[75].

An ongoing formative evaluation of a course need not be a very elaborate procedure. To be sure that the course is "reaching" its audience, however, some feedback should be obtained as early as possible. In the case of, say, a three week programme, there seems little point in waiting until the end of the first week. Indeed, there is much to be said for the use of brief reaction questionnaires completed by the students at the end of each day. The participants should not object to this as long as they know that the data are to be used, as necessary and possible, to improve the course as it progresses. Of course, this requires a commitment from the organizers and teachers that the data will be reviewed daily and that every effort will be made to respond to them.

For ongoing formative evaluation purposes (whether questionnaires are used daily, weekly or at the end of each "unit" in the course), students should be asked their opinions on some or all of the following points:

1. The importance and relevance of the materials covered up to this point. If the course can be divided into a series of discrete sections, an assessment of the relevance of each section would be appropriate.

2. The quality of the teaching in terms of how the material has been presented. If a number of different lecturers are involved, students should be asked to rate the performance of each, particularly if they are to perform again later in the course or if they may be under consideration for use in later programmes.

If a number of different teaching approaches or teaching materials have been used in the course (e.g., films, transparencies, demonstrations, lectures, discussion groups), students should indicate how successful each has been as a means of presenting the information.

3. The novelty of the information presented. Except in the case of a "refresher" course, a student will normally attend an educational offering in order to learn something new. One should therefore be interested in discovering how much the student is really learning (i.e., what proportion of the material he has been exposed to is new to him). Relevance and novelty are both important in the evaluation of teaching. The subject matter of a course may be relevant to a student's interests but not new to him or it may be new to him but not directly relevant to his needs.

4. The "level" of the material presented. It will be important to know, as a course is progressing, whether the material presented is at a level appropriate for the particular audience addressed. It should not be too simple. The students must not feel that they are being "talked down to". Simplicity may also be related to novelty - the students may be learning nothing new because the lecturers have underestimated their prior knowledge of the subject matter and are treating the material at too basic a level. Neither, of course, should the material presented be too complex for the audience (i.e., it should not be "over their heads"). This matter of level of treatment is extremely important in formative evaluation: lecturers must know, early in the course, whether or not they are "reaching" their audience. There is no point in waiting until the end of the course only to discover that the majority of the participants have been completely bored, either because the material has been redundant or because they were unable to follow it.

The problem of "level" is likely to be a critical one in the presentation of workshops, seminars and courses as part of Unesco programmes for the simple reason that participants may vary widely in backgrounds and experience. The fact that they come from different cultural backgrounds and different educational systems, together with the fact that their native tongues may not be the language in which the course is presented, makes the situation unusually difficult. The problem is compounded by the fact that the lecturers will usually not have met the students before the course begins and it may take a little time for them to find out exactly what the students already know about the subject matter.

With this type of heterogeneous audience it may be extremely difficult to conduct a course at a level that will be appropriate to all. It may be necessary to assume, at the outset, a minimum of knowledge on the part of all participants and to use the early part of the course to bring all participants up to about the same level, even if this means that some of them will find this early material redundant.

5. How well the student feels that he is progressing toward satisfying his own objectives in attending the course.

6. An indication of which features of the course, as conducted so far, have been most valuable or interesting and which have been least valuable or interesting.

7. If outside projects or readings have been used in the course, the students' assessment of the value and relevance of these experiences.

8. Any suggestions the students have as to how the course might be changed, in later days, to make it a more valuable educational experience.

9. The overall rating of the course, on some scale, by all participants.

10. Any other observations or comments the students would like to make.

A formative evaluation does not necessarily have to be this highly structured. There may be something to be said for a more free-form approach. Woodward and Yeager[176], for example, used daily logs in which students recorded general impressions of each day in an anecdotal, free-response form. In this they were guided only by very general headings such as "problems encountered".

In all of this data gathering the evaluator must try to discover why various responses are made. If students express dissatisfaction, one must learn the precise nature of this dissatisfaction in order to make changes designed to improve the course. Thus, one must discover exactly why a student feels he is not making satisfactory progress in the course, which parts of it he has found too complex, which parts repetitious, and so on. This implies that the questionnaire should be designed in such a way that, whenever the student indicates some form of dissatisfaction, he is asked to describe the precise nature of the problem.

The kinds of data needed in order to conduct an ongoing formative evaluation of some educational offering are summarized in Exhibit 4. The tabulation is undoubtedly incomplete. In particular, each course or workshop may have some special features that require emphasis in the evaluation questionnaire. Nevertheless, these data seem to be those of greatest overall importance and they should be generally applicable in the formative evaluation of most of the activities with which organizers of courses will be involved. Of course, not all the items listed will apply to every possible situation. Moreover, not everything need be collected at once. A fairly simple daily log (see Exhibit 5) or a check sheet completed after each session (Exhibit 6), for example, could be supplemented by a more detailed weekly review form. Several examples of questionnaires used in formative evaluation are given in Appendix 1. A more detailed discussion on the design of questionnaires is given in section 5.

A terminal evaluation of a course may have formative and summative elements. The sponsors and organizers may need to make decisions concerning the future of the course - whether it will be offered again, with how much change, for what type of audience, and so on - and such concerns can be considered summative. On the other hand, the terminal evaluation may also be used to gather more specific information that may assist the organizers or teachers in improving details of the course if it is to be given again.

In the terminal evaluation of the course one may want to ask some of the same questions that were used in the ongoing evaluation, this time addressed to the course as a whole rather than just a part of it, but some additional types of questions

Exhibit 4

Types of Data to be Collected in the Ongoing Formative Evaluation of a Course

1. Importance and relevance of subject matter.
2. Quality of the presentation
   a) by individual instructor
   b) by form of presentation used.
3. Novelty of the information transmitted.
4. Intellectual "level" of the material presented.
5. Students' assessment of their own progress.
6. Most and least valuable features of the course so far.
7. Value of outside projects and/or readings.
8. Student suggestions for improvement of the course.
9. Overall student rating of the course.
10. Any other observations the student wishes to make.

may also be included. Some of the major questions to be answered are presented in Exhibit 7. Most of these are self-explanatory. At this point it will be necessary to ask students to view the course in its entirety as an educational experience, including various factors that may have contributed to an atmosphere conducive to learning. These factors will include physical arrangements (the lecture and meeting facilities and hotel or other accommodations) and organized social events. For international programmes, one should also discover whether or not the participants were given enough practical information (e.g., on climate, dress, housing) before they left their homes. In this terminal evaluation one is concerned with looking back on the course, to determine how successful it was *in toto* and to identify any failures or problem areas that may have occurred. On the basis of this experience one may be able to make valuable changes to the course in the future or draw on this experience in the planning and execution of future courses in other subject areas.

Bryant[4] has emphasized the importance of several principles of adult education enunciated by Knowles[33]:

(1) Learners must feel the need to learn.

(2) The learning environment is characterized by mutual trust and respect.

(3) The learners must perceive the goals of the learning experience to be theirs.

Exhibit 5

Daily Log

	Excellent	Good	Fair	Poor
Lectures	☐	☐	☐	☐
Demonstrations	☐	☐	☐	☐
Group Discussion	☐	☐	☐	☐
Practical Exercises	☐	☐	☐	☐

Comments/explanations. Use the back of this form to record any comments you would like to make on what has occurred in this course today. We are particularly interested in learning of any problems you have encountered, or if you have been dissatisfied with some aspect of the course. We welcome any suggestions you care to make on how the course might be made more useful for you.

Exhibit 6

Single Session Check Sheet

Reproduced from Stake et al.[64] by permission of the University of Illinois

Name:_____ Date:_____

The following questions are intended to obtain a quick summary of your impressions of the last session.

Please place a check mark (X) to represent your opinion along each scale.

1. How important was the topic for you?

extremely unimportant	neutral	extremely important

2. How helpful was the session to a teacher trying to teach more about the environment?

extremely unhelpful	somewhat helpful somewhat unhelpful	extremely helpful

3. How interesting was the session?

extremely boring	neither, or some of both	extremely interesting

4. If the session were repeated, how would you want to see it changed?

_____

_____

_____

_____

5. Have you any other comments, cri icisms, suggestions?

_____

_____

_____

Exhibit 7

Some Important Questions to be Answered in the Terminal Evaluation of a Course

1. Have students' objectives been met? Have overall course objectives been met?
2. How valuable has the course been to participants in terms of their own professional goals? What were the main benefits gained?
3. How do students rate the quality of the instruction
   a) by individual instructor
   b) by type of learning experience (lecture, discussion, practical sessions, and so on)?
4. Size of group participating--too large, too small, about right?
5. Was the level of treatment of the subject matter "correct" for the majority of the audience?
6. Was coverage of subject matter comprehensive or were there notable gaps in coverage?
7. Were the lectures and/or other educational experiences well integrated into a meaningful and complete programme, or were they disjointed and overlapping?
8. Was most of the material presented "new" to most of the participants, or was it repetitive and redundant?
9. Which features of the course were most valuable and which least valuable to those attending? Another way of putting this is in the form of features that participants would like to add to or delete from the course.
10. Were the physical and related arrangements (lecture halls, accommodations, meals and/or coffee breaks, social events) satisfactory?
11. Would participants recommend this course to their collegaues and would they attend another, similar course if it were offered?
12. Any other ways the course could be improved?

(4) The learners accept a share of the responsibility for planning the learning experience.

(5) The learners participate actively in the learning process.

(6) The learning process is related to and makes use of the experience of the learners.

(7) The learners must have a sense of progress toward their goal.

Bryant suggests that it is these principles that should guide the evaluation of an educational programme, and he puts forward (Exhibit 8) a terminal evaluation instrument designed to determine to what extent these principles have been observed in a particular programme. The principles appear valid and important but many of the evaluation questions that Bryant derives from them are too abstract: phrased in terms that may be meaningful to the specialist in evaluation but too difficult for many students to comprehend or interpret correctly. The opinions of participants may be considered "soft" evaluation data but the questions asked can and should be quite concrete.

One way in which evaluation data can be made more concrete is through focusing attention on the objectives of the students or of the instructors. Instructors must develop a set of behavioural objectives for the course, with one or more behavioural objectives associated with each segment. If, at the conclusion of the course, each student is asked to indicate how far he feels these behavioural objectives have been met, this can legitimately be regarded as part of the evaluation of reaction. If, on the other hand, some "demonstration of achievement" is developed for each of the behavioural objectives, and some objective testing procedures used to determine how far the behavioural objectives have been met, it is really the learning of the students that is being measured and this form of evaluation then falls legitimately in the category "evaluation of learning acquired", as discussed in section 2 of the report.

Meyer and Jenkins[44], in the evaluation of a minicourse, had students use a four-point scale to indicate the extent to which they considered the course objectives were met. The format used is shown in Exhibit 9. The students, presumably, will have objectives of their own. An important element in the terminal evaluation will be the determination of how well the objectives of the students have been satisfied. To do this one must discover, for each student, what his or her objectives were in attending the course, and how far these objectives were actually met. For this purpose, it might well be desirable to use some very brief "pre-course questionnaire" (perhaps incorporated into the actual application form) to determine student objectives before the course begins. At the end of the course, each participant can be presented with the objectives specified before the course began. He is then asked (a) to modify this statement of objectives if he feels that, in retrospect, it was too narrow, too broad, or otherwise inappropriate, and (b) to judge the success of the course in meeting his original or revised objectives. It will also be desirable to include a statement of the objectives of the course as prepared by the organizers and to ask students how far they feel that these general objectives have been met.

The California State Department of Education[16] recommends that, at the beginning of the programme, each student be asked to prepare his own set of goals. As the course progresses, he is invited to modify this list and the terminal evaluation should include some questions of the following type:

Exhibit 8

Evaluation Instrument Proposed by Bryant[14].

Reproduced from The Evaluation of Continuing Education for Professionals: A Systems View, eds., Le Breton, Bryant, Zweizig, Middaugh, Bryant, Corbett (copyright © 1979 by the University of Washington, Seattle, Washington 98195) by permission of the University of Washington.

Instructor Evaluation

Name of Instructor _____

Name of CE Program _____

Instructions: Please place an X at that point on the continuum that best represents your answer.

		Not at all	Very Much
(1)	To what extent did the instructor expose you to new possibilities for self-growth?	0	5
(2)	To what extent did the instructor help you to clarify your desire to improve your skills?	0	5
(3)	To what extent did the instructor aid you in the diagnosis of the gaps between your aspirations and your present level of performance?	0	5
(4)	To what extent did the instructor help you respect your own feelings and ideas?	0	5
(5)	To what extent did you feel mutual trust and helpfulness among students?	0	5
(6)	To what extent did you feel a spirit of mutual inquiry between yourself and the instructor?	0	5
(7)	To what extent did you feel there was a mutual process of setting learning objectives?	0	5
(8)	To what extent were you able to share your thinking about options in designing learning experiences, selection of materials, and methods of instruction?	0	5
(9)	To what extent did the instructor help you to organize a learning-teaching environment in which the responsibility for the process of inquiry (learning-teaching teams, task groups, independent study, etc.) was shared?	0	5
(10)	To what extent did the instructor draw on your own experiences as resources for one anothers' learning?	0	5
(11)	To what extent did the instructor gear the presentations to your level of experience?	0	5
(12)	To what extent did the instructor involve you in mutually acceptable criteria for measuring your progress toward learning objectives?	0	5

Exhibit 9

Student Evaluation of the Extent to Which Course Objectives Were Met

Reproduced from Meyer and Jenkins[44]
by permission of the Association of Educational and Training Technology

Objective	Rating			
	0	1	2	3
At the conclusion of this minicourse you should be able to:				
1. identify the steps involved in planning and implementing a curriculum project within a school;				
2. work effectively as a member of a curriculum development team within a school;				
3. describe and discuss various factors influencing the determination of aims and objectives for various levels of the curriculum within a school;				
4. apply strategies for: (i) collecting information (ii) making effective decisions (iii) improving communications both vertically and horizontally within the school administrative structure in relation to curriculum development;				
5. produce a practicable plan of action for improving the process of curriculum development in your school				

Rating Key : Extent to which each objective has been attained.

3 = Sufficiently for all practical purposes
2 = Adequately, but some room for improvement
1 = To a minimum extent only
0 = Not at all

1. What were your goals when you entered the programme?

2. Have you changed your mind about any of these goals? If so, how?

3. Which of your goals have been achieved?

4. If you did not achieve some goals, what could have been done to help you accomplish them?

5. What do you intend to do with your knowledge, skills and capabilities; that is, what are your goals for the future?

In evaluating an institute for the training of librarians, held in the Caribbean, Chang[8] approached this situation through a very simple pre-test and post-test. The pre-test instrument posed the following question?

> Please list specifically what you expect to learn from attending the Institute for Training in Librarianship.

The corresponding post-test question was put as follows:

> On your pre-tests, you were asked to list specifically what you expected to learn from attending the Institute. Below, please indicate whether or not your expectations have been met by the Institute. Please give examples. If your expectations have not been met, please indicate in what way(s) they have not been met.

Note that Chang seeks to discover the causes of failure or dissatisfaction. As emphasized earlier, it is important to discover how and why a failure occurs if such failures are to be corrected or avoided in the future.

A summative evalution may not be entirely structured. It is sometimes illuminating, for example, to have each student list the things that they themselves feel they have learned as a result of their participation.

Patton[48] strongly advocates the use of in-depth, open-ended interviews for a variety of evaluation purposes. He gives an example of a series of questions used in the terminal evaluation of a 10-day course involving a group of disabled and able-bodied participants, as follows:

1. To what extent was the course what you expected it to be?

   a. How was it different from what you expected?
   b. To what extent did the things you were concerned about before the course come true?
      b-1. Which things came true?
      b-2. Which didn't come true?

2. How did the course affect you personally?

   a. What changes in yourself do you see or feel as a result of the course?
   b. What would you say you got out of the experience?

3. During the last nine days you've been with the same group of people constantly. What kind of feelings do you have about having been a part of the same group for that time?

   a. What feelings do you have about the group?
   b. What role do you feel you played in the group?
   c. How was your experience with this group different from your experiences with other groups?
   d. How did the group affect you?
   e. How did you affect the group?
   f. In what ways did you relate differently to the able-bodied and disabled people in your group?

4. What is it about the course that makes it have the effects it has? What happens on the course that makes a difference?

   a. What do you see as the important parts of the course, that make an Outward Bound course what it is?
   b. What was the high point of the course for you?
   c. What was the low point?

5. How do you think this course will affect you when you return to your home?

   a. Which of the things you experienced this week will carry over to your normal life?
   b. What plans do you have to change anything or do anything differently as a result of this course?

There is little doubt that, in the hands of a skilled interviewer, such open-ended questions can produce valuable insights. A major problem of this approach is the accurate recording of responses; it will probably be necessary to use a tape recorder during interviews. Great skill will also be needed to reduce the interviews into a meaningful summary for reporting purposes.

It should be pointed out that some writers are critical of the kind of terminal evaluation typified by such questions as "How much of the material was new to you?". Misanchuk[45], for example, refers to this as justificative evaluation (an attempt to justify the event after it has been completed) and claims that this type of evaluation should not take place at all. He states that such evaluation asks participants for information that should have been gathered prior to the design of the course, as part of the "needs assessment" or in the selection of participants. To a certain extent he is correct of course. Nevertheless, he is referring to a rather ideal situation. In the case of courses involving international students, which will frequently be true of Unesco offerings, it may be extremely difficult to gather reliable information concerning the backgrounds of students before the course is actually underway.

Both Misanchuk[45] and Dixon[12] indicate that the subjective opinions of students may not always correlate well with objective measures of learning taking place (i.e., the student may express greater satisfaction than he really should). It is possible that one should only give opinion great weight if it agrees well with other measures. Dixon, in fact, suggests that the opinions and perceptions of students may have more validity if collected later, perhaps six months or a

year after participation in the programme, so that the students have had some opportunity to apply what they learned. In evaluating a series of workshops, Jaster[29] supplemented an evaluation form used immediately after the workshops with telephone interviews conducted with the participants one or two years later. His use of this technique is discussed further in section 3.

1.2. Independent observers

So far, only the reaction of students has been discussed. While student reaction is obviously of great importance, it is not the only data of interest. Wherever it can be arranged, it may be highly desirable to have some independent observer present at some or all of the presentations. Such an observer can serve two functions: (1) to mix with students and obtain feedback and reaction that may be more "candid" and impromptu than the reaction data obtained by the teaching staff through more formal procedures, and (2) to observe the teachers in action and thus to arrive at his own assessment of the quality of the course, both in terms of its content and in terms of the methods of presentation. This second function may be regarded as a form of "peer review". It requires, therefore, the use of an individual who is knowledgeable in the subject-matter covered in the training programme and at approximately the same level of seniority and professional experience as the members of the teaching staff themselves. It is probably desirable to provide this outside observer with some type of checklist on which to record, in standardized form, observations on course content, teaching methods, student interest, and whatever else one may be concerned about. An example of such a checklist is given in Appendix 2. On the other hand, an observer may be encouraged to make less structured diary-type observations as illustrated in Exhibit 10.

Exhibit 10

Excerpt from the Diary of an Evaluation Observer

Reproduced from A. D. Grotelueschen et al. Evaluation in Adult Basic Education.[24] Copyright 1976, Interstate Printers and Publishers Inc. Reproduced by permission of the publisher.

May 18, 1975

Today we completed the last of our observations in the ABE program. Some rather interesting things have emerged in these observations which we'd like to try to summarize.

There appears to be some contradiction, or at least discrepancy, between what the teachers say they think the ABE program ought to be about and what actually seems to be happening as we were able to observe it.

For example, in our interviews with teachers, most said they felt the program was so oriented toward reading that other important things were being slighted. The most common thing teachers referred to was the whole idea of coping skills or the things people need to know how to do to get along in the world. Teachers felt these skills were essential and said it was likely that the ABE program was the only place these adults were able to learn these skills.

In practice, however, there is little evidence to suggest that these same teachers are able to do much with teaching coping skills. Classroom observation indicates that up to 95% of instructional time is spent in reading or language skills areas.

Houge[27] has described the use of "participant observers" in evaluation. A participant observer is a bone fide registrant in a course who, before the course begins, agrees to act as a kind of official observer of what transpires. Such an observer will become familiar with the goals and objectives of the programme, will meet with the faculty, make observations on teaching techniques, and interact with the other students in order to collect their perceptions. Houge claims that this method complements and is consistent with more conventional methods of getting participant reaction. Patton[48] points out that participant observation is essentially a combination of observing and informal interviews.

Observation can be a powerful evaluation method if properly used. Patton[48], for example, has claimed:

> To understand fully the complexities of many program situations, direct participation in and observation of the program may be the best evaluation method. (Page 30).

His book presents an excellent account of evaluation through observation, including the use of participant observers.

1.3. Reaction of instructors

Another "reaction study" will involve the teaching staff. It will be valuable to have the lecturers evaluate the course from their point of view, and to prepare a summary of this evaluation once the course is concluded. A good lecturer will constantly be evaluating himself. He is likely to recognize the fact that certain material was not presented as clearly or as completely as he would have liked, that a different sequence of presentation would have improved the situation, that certain types of information could be better presented in an alternative way (e.g., by student readings and group discussion instead of straight lecturing), and that certain topics might be omitted entirely in the future, because they were redundant, of only marginal relevance or interest, or because they were clearly boring to the audience. The lecturers' evaluation should also include their observations on the students - their quality, degree of interest in the subject matter, the intelligence of the questions asked, their diligence (e.g., in completing background reading and other assignments), and their general suitability for participating in a course of this type. It is possible that a course may fail to meet part of its objectives because some of the students did not have the background necessary to benefit fully from the experience. This could be a significant problem in the case of a programme that may be attended by participants from many different countries, with widely divergent backgrounds, and varying levels of competence in the language in which the course is offered. In any case, the lecturers' evaluation of the course, and of the student body, is likely to be an important element in a complete evaluation programme, and such an evaluation may have great significance in the advertising of future courses and in establishing appropriate criteria for the selection of candidates. The teaching staff should also evaluate the "environment" of the course: the adequacy of the facilities, the helpfulness of the local organizers, and other factors that may add to or detract from the educational aspects.

Knowles[33] has mentioned that instructors have great shortcomings as observers because they

> . . . are personally involved in the outcome of the evaluation, so that it may be difficult for them to be objective. They may tend to overlook instances in which desired changes are not being produced and to emphasize minor successes. (Page 237).

On the other hand, it could be argued that the reverse situation might also be true. Some instructors may be too sensitive, too self-critical. They may find faults where none really exist. Whatever the limitations of the instructor as an evaluator, however, it is clear that he has an important role to play. His input will be analysed and interpreted along with the input from students, independent observers and other individuals who are involved in some way with the educational experience. It is important that a complete evaluation should be based on input from a number of individuals representing different levels of involvement and points of view.

In the case of educational programmes in which many different instructors are involved, each one called upon to cover a particular segment, it may be valuable to obtain some reaction data from each. A form for this purpose, as used by Slater[58] is shown in Exhibit 11. Note that this gathers opinions on a variety of concerns: the organization and planning of the programme, the facilities, the audience, and the instructor's perception of his own performance.

Exhibit 11
Instructor Evaluation Form
Reproduced from Slater[58] by permission of Aslib

(1) How do you feel your lecture (or demonstration, practical, forum, etc) went over?

_____ well

_____ satisfactorily, average

_____ unsatisfactorily, not so well as hoped

(2) If your answer was "unsatisfactorily"–why would you say that was?

(3) From what you saw of them, would you say attenders on this course were:

_____ a good audience

_____ average

_____ below average

_____ a very mixed audience

(4) Were they the kind of people you expected? (In terms of background, experience, qualifications, status etc).

_____ much as expected

_____ some were, some were not; a mixture

_____ on the whole not as expected

(5) Was the basic aim of your contribution to the course:

_____ to give an introduction to a subject

_____ to provide further education in a subject; background information; broaden knowledge of audience

_____ to indicate range and nature of sources available in a subject

_____ practical guidance in how to use sources and services

_____ practical instruction in specific techniques, or in use of equipment

_____ something else (please describe)_____

_____

(6) What about the advance notice and briefing you were given for your lecture or demonstration (etc.) how was this–

(a) in terms of time?

_____ notified too long in advance really

_____ good notice, plenty of time to prepare, just right

_____ adequate notice

_____ not quite enough time

_____ really insufficient time, inadequate notice

(b) in terms of subject content and scope: what you expected to say or do?

_____ very good, couldn't have been much better

_____ good

_____ adequate

_____ slightly inadequate

_____ really very inadequate

(7) Ideally, what kind of brief do you like to get? (in terms of time, subject content and scope, and any other relevant factors)

Omit this question if you have answered it before on forms distributed by this organization.

(8) How much did you know about what other lecturers, demonstrators, panel members (etc.) were going to say, show, or do, and the line they were taking?

_____ quite a lot

_____ enough to cope

_____ not enough really

(9) Have you any comments you'd like to make on the facilities available to you as a lecturer or demonstrator (etc.) on this course? (Is there any way things could have been made easier or more convenient for you?)

(10) Do you feel that this course as a whole:

_____ is good, in need of no real improvement or change at the moment

_____ is adequate, meets a need, but could be improved in minor ways

_____ is unsatisfactory really, changes or replanning seem definitely necessary

_____ I don't feel sufficiently well informed about the rest of the course to give an opinion

(11) If you feel that any changes (major or minor) are needed, or would help this course, what would you suggest?

2. EVALUATION OF LEARNING

While the opinions of programme participants can be considered an important element in evaluation, more objective measures of the success of an educational programme may be desired. In particular, the organizers of a course may want a more precise measure of how much the participants have learned. Clearly, this is more difficult than merely gauging student reaction.

The use of behavioural objectives was alluded to in Section 1. An instructor should develop behavioural objectives for each segment of a course and, for each objective, identify an appropriate "demonstration of achievement". How precise these objectives and demonstrations can be depends largely on what is being taught. Psychomotor skills (e.g., how to assemble a piece of equipment) can usually be reduced to rather precise objectives and demonstrations. Cognitive objectives may be quite precise if they relate to the imparting of factual knowledge but will tend to be much less so if they relate to the improvement of analytical and problem solving skills. Affective objectives, having to do with changing the attitudes or opinions of a group, tend to be least easily converted to concrete demonstrations of achievement.

In developing behavioural objectives one can be guided by much sound advice in the literature of education (see Wilsing[74] for a recent example). To the extent that an objective can be made precise and converted into an immediate demonstration of achievement, it may be possible to test a student's learning during the course itself. For example, a possible objective might be:

> To teach students how to construct a thesaurus conforming to the guidelines for thesaurus construction approved by the International Organization for Standardization

and an appropriate demonstration of achievement might be:

> The student constructs a mini-thesaurus of approximately 100 terms showing the correct use of the BT, NT and RT relationships, use references and scope notes. The BT, NT, and use relationships must be properly reciprocated.

In some cases, however, the objective and demonstration will be of a longer-term character, as in the following example:

> By September 1983, 90% of workshop participants will have conducted workshops in their own states, interpreting the knowledge gained here.

Clearly, the extent to which this objective is reached could only be determined through some follow-up procedure.

If the main objective of an educational programme is to produce some attitudinal change among participants, it may be possible to assess the success through some form of pre-test and post-test of attitudes. A series of attitudinal statements makes up the test. Respondents indicate their agreement with each statement on a scale such as: Strongly Agree, Agree, Disagree, and Strongly Disagree. Such a method has been used by Smith[59], Penn[50] and Postlethwait et al[51], among others. Postlethwait et al. used such statements as:

I approach biology with a feeling of hesitation.
I really like biology.
I have always enjoyed studying biology in school.
It makes me nervous to even think about doing a biology experiment.
I feel at ease in biology and like it very much.
I feel a definite positive reaction to biology; it's enjoyable.

in evaluating an undergraduate course in biology, and Penn employed statements of the following type:

The issue of sex role stereotyping is equally important for both males and females.

The concept of career should include both work and family.

I am committed to continually monitoring my own sex role stereotyping attitudes and behaviours.

to assess attitudes toward sex role stereotyping.

The measure of success, then, would be the extent of movement in student attitudes (before and after the course) toward the desired objectives.

A more elaborate type of evaluation would involve the use of a control group. The attitudes of students who had taken the course would then be compared with the attitudes in a matched control group (see, for example, Mehlinger and Patrick[43]).

In a few cases, of course, a desired change in attitude or behaviour may be directly observable by the instructors. This might be true, for example, in a course on public speaking or one on "leadership".

Silver[57] has warned that the attitudes of participants toward the subject matter of a course may influence their opinions on the quality of the course. He describes an experiment in which, before a course, the knowledge and attitudes of participants were measured. After the course, tests of knowledge and of attitudes, as well as a questionnaire relating to the quality of the course, were applied. Silver found that attitudes did not change significantly before and after the course and were found to be directly related to opinions on the quality of the course. That is, those participants with the more positive attitudes towards the subject of the course were more likely to judge it of high quality. On the other hand, opinions on quality did not seem to correlate with changes in the participants' knowledge of the subject matter.

In the case of courses designed to impart factual knowledge, some test of mastery of subject matter can be applied to students before and after the course. It is highly desirable that the test be devised in a form suitable for true/false or similar check-off responses. As an illustration, some sample questions used to test a student's knowledge of energy-related matters is shown in Exhibit 12.

In a pre/post situation involving factual-type questions, there are several possibilities for bias in one direction or another. If the teaching staff prepare the set of questions there is a danger that, consciously or unconsciously, they will give particular emphasis to these questions in their actual presentations. If the same questions are then used at the end of the course, the "evaluation of learning" may be somewhat biased in favour of the course since it is possible that other questions, although equally important, might not be answered nearly as well by the group of students. It is also possible that the students themselves, remembering

Exhibit 12

Example of Statements for Testing Factual Knowledge

Reproduced from Stake et al[64]
by permission of the University of Illinois

(1) The population of the United States doubles about every 46 years and the demand for electricity doubles about every 10 years.

(2) Widespread conversion of urban refuse into fuels could supply approximately 25 percent of our total annual energy consumption.

(3) If all coal from reserves in the U.S. were mined and made available, there would be enough to last us for several hundred years if used at the present rate of consumption.

(4) The cost of transporting coal can often equal the cost of mining.

(5) Oil is the cleanest burning of the fossil fuels and is, therefore, in great demand.

(6) The most significant variable affecting fuel consumption in an automobile is its weight.

(7) The best location for the thermostat is on the coldest wall.

(8) Fluorescent lights are less efficient than incandescent.

(9) The energy used to drive a car initially came from the sun via green plants.

(10) The heat lost through the house is primarily through the ceiling.

(11) In the Tennessee Valley region, the water heater accounts for about 50 percent of the electric bill.

(12) The electric range uses less electricity than the clothes dryer, window air conditioner and diswasher combined.

the questions asked in the pre-course questionnaire, will concentrate on these topics in their study, possibly to the exclusion of other topics of equal or greater importance. This could be true even when the student is given no indication that he will be tested again at the end of the course, a condition essential to the conduct of an evaluation of this type.

There are ways in which some biases can be reduced (e.g., the questions may be compiled by an independent evaluator or use may be made of a crossover test, in which students and questions are divided into two groups so that group A answers A questions before the course and B questions after the course, while group B answers B questions before the course and A questions afterwards) but these may introduce new biases.

It can be seen, then, that it is somewhat difficult to arrive at a test design that has no possibility of bias in one direction or another. From the viewpoint of

experimental design, it would be better to use a control group of people who have not taken the course but are otherwise well "matched" with the student group in terms of other characteristics, especially in their educational background and level of experience. The use of a control group in this way would eliminate the need for the use of pre-course questions. The two groups would simply answer the same set of questions when the course is completed, and the two sets of results can then be directly compared, the assumption being that the student group should get significantly better scores than the control group. This type of measurement of learning, although possible within a university environment, would be extremely difficult to implement with most courses, where potential groups of participants are much less "captive".

Perhaps the best compromise is to develop a rather large set of questions that range over the entire subject matter of the course. One half of these questions are then selected at random to form the pre-course test and the other half form the post-course test. The "at random" is important here. As Braskamp et al./3 have pointed out, an instructor could achieve very good results by choosing difficult questions for the pre-test and easy ones for the post-test.

It is well to recognize, as Linn/40 discusses, that the difference between a pre and post score for an individual is typically an unreliable indicator of actual change in knowledge. The reliability improves considerably, however, when scores for an entire group of individuals (e.g., all participants in a course) are considered in the aggregate.

Mehlinger and Patrick/43 point out that an evaluation instrument that tries to measure learning must satisfy three basic requirements to be valid: test items must be specifically related to the stated objectives of the course, experts must be able to agree on what is the "correct" or "best" answer to a question, and most students who have not taken the course should not be able to respond correctly to the test items.

Veri et al./72 refer to requirements for the evaluation of learning in the following terms:

> An ideal procedure is to measure the cognitive, affective and psychomotor skills of the learners before they enter the learning process, teach to behaviorally defined objectives, and then measure what has been learned with somewhat detailed objective tests. The tests must be designed meticulously using accepted test construction methods and procedures. Follow-up testing at some later time to measure retention of learning completes the ideal model. (Page 4).

Clearly, it is easier to test learning acquired in a course if its main purpose is to impart factual information or if it is designed to teach a particular skill (e.g., the production of a thesaurus). But a course may be designed not to impart factual information but to deal with problem solving or decision making situations. It is more difficult to test the impact of a course on the problem solving or decision making skills of participants but it is not impossible. One possible technique has been described by McGuire and Babbott/41. The test was devised to assess the abilities of medical students in taking care of a sick person. Each student is given a medical case history on the patient, along with several possible diagnoses or courses of action. The student must choose among the alternatives presented and each choice branches to other choices. Eventually, the patient will either recover, be referred elsewhere, or will die. A panel of medical experts is used to rate each

choice at each decision point on a five-point scale, ranging from "clearly indicated" to "clearly contraindicated". An efficiency score, a proficiency score and a composite score are calculated for each student. The efficiency score represents the proportion of the student's responses that are beneficial to the patient and the proficiency score represents the percentage agreement between the student's choices and those of the panel of experts. The composite score combines the efficiency and proficiency scores. Although this test was designed for the health care situation, it is clear that the techniques could be modified to apply to other types of decision making environments.

3. EVALUATION OF BEHAVIOURAL CHANGE

If evaluation of learning is difficult, evaluation of behavioural change in students is even more of a problem. This phase of evaluation is concerned with the long-term effects of an educational programme. This goes beyond learning as such into the application of the learning acquired from a particular course. The obvious consideration in this evaluation is to find out how the students have benefited, in the long run, from some programme. Benefits might relate to the professional advancement of the student: promotion, increased responsibility, successful movement into a different type of position (e.g., into teaching or research), and so on. Hampton[26] suggests that one way of accomplishing the evaluation of behavioural change is by the "systematic appraisal of on-the-job performance before and after the educational experience" and that such an appraisal should be made "by the person receiving the education, his superiors, his peers, and his subordinates".

It seems quite likely that the organizers or sponsors of courses will be interested in this level of evaluation for some at least of their educational offerings. A Unesco course, for example, may have as a short-term objective the training of teachers, and prospective teachers, in some subject areas. The longer-term goal may be the further dissemination of the knowledge acquired in the course. This goal would be achieved if the course participants were able to use some or all of the course material in their own teaching, or in the planning of curricula, when they return to their own countries. It is clear that the long-term success of this course could only be measured by some type of follow-up study, conducted perhaps six months to a year after the end of the course. A follow-up questionnaire can be used in this situation to determine whether or not the course participants have been able to make use of the material presented in their own teaching and, if they have used it, with what degree of success. At the same time, the participants can be asked to look back on the course and to assess its value within the longer-term perspective.

With the kinds of individuals and situations dealt with in many international settings, it may not be appropriate to attempt formal on-the-job performance assessments, using peers or superiors of those participating in the programmes. Nevertheless, there may be situations in which it will be possible and appropriate to conduct some form of follow-up to gain a further and different view of the results of a programme. For example, some agency of a national government may nominate or sponsor one or more students to participate in a particular course. It may be desirable and feasible to ask this agency, some months after the completion of the course, to provide feedback on the progress and performance of these individuals. In this way, one can get some idea of how satisfied the various national governments are with the programmes developed and offered.

For most courses, it will be worthwhile to at least conduct a follow-up with the students. This follow-up, by means of a questionnaire, will be conducted to obtain each student's longer-term view on the value of the course, its impact on his own professional development, and the extent to which he has been able to apply the knowledge acquired.

Nadler[46] contends that evaluations performed some time after the conclusion of a course are likely to achieve more reliable results. Because of "psychological disengagement" (participants are thinking more about their travel home) or a "halo effect" (e.g., the latest event in the course may exert most influence even if it

is atypical of the course as a whole), the evaluation data provided at the immediate conclusion of a course may be superficial and misleading. At the same time, Nadler recognizes the problems involved in getting participants to return questionnaires some months later.

Nadler suggests that, before the conclusion of the course, all participants should be asked to write a brief memorandum indicating specifically what they intend to do as a result of participating. Three copies are made: one retained by the student and two by the evaluator. Some months later, the evaluator mails one of these copies to the student to provide "linkage" to the course. While not specifically suggested by Nadler, this occasion could be used to ask the recipient to what extent the planned actions have actually taken place.

Nadler recommends various approaches to maintaining "linkage", designed to ensure that a course has some continued impact on participants. One such technique is the development of an informal newsletter in which participants share experiences on what they have done as a result of attending.

Slater[58] includes the following items in a follow-up questionnaire used in the evaluation of short courses in the field of library science:

(1) What is your assessment of this course as a whole, now:

..... good

..... average, adequate

..... disappointing

(2) If this course were repeated, and a colleague asked your advice about whether it was worth attending, what would you say

..... yes, go

..... no, don't go

..... it depends (on what? Please explain briefly below)

(3) (a) Looking back on it, does anything about this course stand out as being particularly helpful to you professionally?

..... yes

..... no

(b) If "Yes" above: What is it?

(4) Have you actually used, or since put into practice, anything new you heard about at this course?

..... yes

..... no

(5) If "Yes" above: What was it - and how did you use it?

(6) If "No": Was this because:

..... you learned nothing new at the course

..... you heard about new things, but they are not appropriate to your situation, and you are not likely to use them

..... you heard about new things, but haven't had a chance to use them yet.

(7) Useful contact with other librarians is sometimes mentioned as a side-benefit of attending courses:

(a) Did you make what seemed to be useful or interesting contacts at the time?

..... yes

..... no

(b) Have they in fact been useful to you since?

..... yes

..... no

(8) What would you now say were the main faults or defects of this course?

(9) If you were put in charge of re-planning this course, what changes would you make to remedy this?

(10) How well do you remember this course?

..... clearly, well

..... reasonably, fairly well

..... not very well

While one might differ somewhat from Slater in question format or wording, these do seem to represent the kinds of questions for which answers should be sought in some follow-up procedure. A simpler follow-up questionnaire, but having the same general objectives, is shown in Exhibit 13.

Jaster[29] used follow-up telephone interviews as a way of assessing the impact of a seminar on written communication skills. The interview schedule followed is given in Appendix 3. Telephone follow-up has much to commend it (not least of which is the fact that it establishes another form of the linkage that Nadler advocates), although it may not always be easy to implement where international audiences are involved. Jaster's interview guide is put forward, not only as an excellent example of the interview format, but because the approach could be modified for use with a mailed questionnaire.

Patton[48] is a strong believer in the use of interviews with participants - before the course, at the end of the course, and as a means of follow-up. An example

Exhibit 13

Follow-up Questionnaire

Reproduced from A. D. Groteluesch en et al. Evaluation in Adult Basic Education.[24] Copyright 1976, Interstate Printers and Publishers Inc. Reproduced by permission of the publisher.

Your responses to the following statements will assist in evaluating the type and extent of impact your participation in the recent continuing education experience has had on aspects of your basic education program.

1. To what extent have you used what you have gained in the continuing education activity?

1	2	3	4	5
Substantially		Moderately		Not at all

(If there has been some application, please answer questions 2 and 3. If not, skip to question 4.)

2. Briefly describe your use of what you gained.

_____

_____

3. Briefly indicate any outcomes associated with this use (e.g., improved instruction or student learning, more paper work).

_____

_____

(Please skip to question 5.)

4. Indicate those factors in the situation or within you that have prevented application.

_____

_____

_____

5. Since participating in the continuing education activity, have you discussed with your colleagues various aspects of the activity?

Yes _____ No _____

6. *(If yes)* How were your ideas received?

Favorably ___ Neutrally ___ Unfavorable ___

7. In general, how adaptable (or applicable) has the material and/or procedures covered in the continuing education activity been for your program responsibility?

1	2	3	4	5
Very	Quite	Somewhat	Hardly	Not at all

of a follow-up interview is given in Exhibit 14. Note how the interview attempts to get reaction directly related to the course and its impact (including behavioural changes resulting from the course) and also probes for attitudinal changes (the participants were interviewed before the course so pre-course and post-course responses can be compared).

Another excellent example of a follow-up questionnaire is given in Appendix 4. This is intended for use by mail and focuses on the most recent training course (in the legal field) attended by the participant. Again, this questionnaire could be modified to meet most follow-up requirements.

As suggested in section 0.2, it would be desirable to get indications of behavioural change by methods other than asking the programme participants. In particular, one would like more objective or even unobtrusive approaches. There may be a few situations when unobtrusive measurements of impact may be possible (e.g., new courses initiated by teachers in their own establishments). In the majority of cases, however, the best that could be done would be to consult people other than the participants themselves - their professional colleagues or supervisors - on, for example, improvements in performance, increased responsibilities, or new projects undertaken, where such activities might stem, at least in part, from participation in the training programme.

In other cases, the long-term objective of a course will be to achieve some effect on a population other than the participants themselves. Consider, for example, a course designed to inform agricultural extension agents of a new method for irrigation. The long-term objective would be to have the improved irrigation method adopted in the farming communities with which these agents deal. In such situations, if one is serious about evaluating the impact of the programme, every effort should be made to use follow-up procedures to determine to what extent the desired outcome (in this case, adoption of the irrigation procedure) has been achieved.

Exhibit 14

Example of Interview With Participants Six Months After Conclusion of a Course

Reproduced from M. Q. Patton. Qualitative Evaluation Methods./48 Copyright, 1980, Sage Publications, Inc.
Reproduced by permission of the publisher.

FOR DISABLED

(*Check previous responses before interview. *If person's attitude appears to have changed, ask if they perceive a change in attitude.)*

4. We asked you before the course to tell us what it's like to be disabled. What are your feelings about what it's like to be disabled now?
   a. How does your disability affect the types of activities you engage in?

      (Clarification): What are some of the things you don't do because you're disabled?

   b. How does your disability affect the kinds of people you associate with?

      (Clarification): Some people find that their disability means they associate mainly with other disabled persons. Other people with disabilities find that their disability in no way limits their contacts with people. What has been your experience?

   c. As a result of your participation in Outward Bound, how do you believe you've changed the way you handle your disability?

5. About half of the people on the course were disabled people and about half were people without disabilities. To what extent did you find yourself acting differently with disabled people compared to the way you acted with able-bodied participants?

6. Before this course we asked you how you typically face new situations. For example, some people kind of like to jump into new situations even if some risks are involved. Other people are more cautious, etc. How would you describe yourself along these lines right now?
   a. To what extent, if at all, has the way you have approached new situations since the course been a result of your Outward Bound experience?

FOR ABLE-BODIED

4. We asked you before the course to tell us what it's like to work with the disabled. What are your feelings about what it's like to work with the disabled now?
   a. What do you personally feel you get out of working with disabled persons?
   b. In what ways do you find yourself being different from your usual self when you are with disabled people?
   c. As you think about your participation in the course, what particular feelings do you have about having been part of a course with disabled people?

7. Have there been any ways in which the Outward Bound course affected you that we haven't discussed?
   (If Yes): How? Would you elaborate on that?
   a. What things that you experienced during that week carried over to your life since the course?
   b. What plans have you made, if any, to change anything or do anything differently as a result of the course?

8. Suppose you were being asked by a government agency whether or not they should support a course like this. What would you say?
   a. Who shouldn't take a course like this?

9. Okay, you've been very helpful. Any other thoughts or feelings you might share with us to help us understand your reactions to the course and how it affected you?
   a. Anything at all you'd like to add?

4. EVALUATION OF PROGRAMME RESULTS

The "evaluation of programme results", as the expression is defined for the purpose of this report, is concerned with the assessment of a complete educational programme, which may involve a whole series of courses or other educational endeavours. The organization responsible for planning and presenting the programme should be interested in discovering how successful the programme has been as a complete entity. This is a level of evaluation that is of broader scope than the type of evaluation considered so far in this report. As an example, one can evaluate as distinct units the several courses that comprise a complete curriculum in an academic department. It is conceivable that each course could be quite successful but that the entire programme fails to achieve its objectives, possibly because it is incomplete or because it emphasizes the wrong things. The programme may be successful in imparting basic technical skills and thus equipping the students to perform adequately in beginning professional positions. However, it may fail to provide the professional philosophy or theoretical underpinnings that will allow the student to "grow" in his profession and thus be capable of accepting increased managerial responsibility. If the objective of the programme is merely to instruct in basic skills, it can be considered successful. If, on the other hand, the programme has the longer-range objective of equipping the student to develop professionally, to adapt to change, and to continue his learning, it may have failed completely. It is quite clear from this that there must be programme objectives as well as course objectives, and that criteria and procedures should be developed whereby the results of the programme can be evaluated against the programme objectives. Evaluation of the programme is the responsibility of the programme planner rather than the individual instructors, although these individuals, along with the students, may clearly have an important role to play in the assessment of the programme.

A series of general objectives should be developed for any educational programme. For any specific educational activity within this programme it will be important that precise objectives, relating to the overall programme objectives, be developed. These objectives must clearly deal with who is to be trained, what information is to be imparted, and what the end results of the programmes are intended to be. The educational activities themselves must obviously be planned with these objectives clearly in mind. In relation to the stated objectives a number of important decisions must be made, decisions concerning how the objectives can be met as efficiently and economically as possible. Questions to be considered will include the following:

(1) What type of educational offering (short course, long course, workshop, seminars, or various combinations of these) is likely to be most appropriate to satisfy the programme requirements?

(2) How many of these are needed to satisfy the programme objectives and how frequently must they be presented?

(3) Where should the training programmes be held in order to have the maximum possible impact at the least cost?

(4) What individuals, types of individuals, or groups are best qualified to implement a particular programme?

(5) What type of student should be admitted to the programme, how many students should be admitted, and how should the students be selected?

(6) What language or languages should the programme be held in?

The above list is intended only as a sample of the types of questions that must be considered in the planning of training activities. The important thing to recognize here is that evaluation must be integrated into these activities from the very beginning and that it must occur at various stages in the complete process. The following sequence is recommended:

(1) Establish objectives for a particular programme;

(2) Evaluate the objectives. Are they really the objectives one wishes to obtain? Are they reasonable and viable? Modify the objectives if they fail to stand up to this examination;

(3) Develop plans for a training programme that is likely to satisfy these objectives as efficiently as possible, addressing all of the various questions mentioned earlier;

(4) Develop criteria by which the programme can be evaluated in terms of the achievement of its objectives;

(5) Develop procedures for programme activities;

(6) Have each element in the programme evaluated. By "element" we mean individual workshop, seminar, course, or whatever other means are used to give the education or training needed. Evaluation of a programme element could include reaction evaluation (formative and summative), evaluation of learning acquired, and evaluation of behavioural changes in the students;

(7) Analyse and interpret the results of these evaluation activities. This should be a continuing and current activity. Programmes may be slightly modified, expanded, drastically changed or abandoned completely on the basis of results gathered through these evaluation activities. It is, of course, important, that the evaluation data be considered in relation to the criteria developed earlier; i.e., the criteria identified as significant in measuring the degree to which the programme objectives have been met;

(8) On the basis of all the evaluation data available, consider whether or not the programme objectives have been satisfied, conducting whatever further surveys (e.g., follow-up studies with students, employers, sponsoring agencies) may be needed to complete the overall evaluation of programme results;

(9) Identify weaknesses or failures and the causes of these. Use the knowledge thus gained in the planning of future, improved programmes.

As Steele[65] has pointed out, programme evaluation goes beyond instructional evaluation. It is concerned with the additive effects of a whole series of instructional units. Instructional evaluation deals primarily with impact on the individual whereas programme evaluation deals more with impact on some group or community of individuals. Programme evaluation is concerned, among other things, with the establishment of priorities.

In this section of the report various broad levels of evaluation applicable to educational programmes have been identified. An example of a very thorough and multi-faceted approach to the evaluation of a training course is given in a report from the U.S. Civil Service Commission[71].

In this course, the behavioural objectives were carefully specified and the degree of attainment of each objective was determined by pre- and post-testing, by post-assessment by each student and by the student's supervisor, and by "observations" made by peers and the teaching staff using a checklist. The pre- and post-test in this evaluation made use of two test sets of fifty questions each, approximately matched as to difficulty. All of these questions were of the true/false or multiple choice type. An interesting technique employed in this evaluation was the use of a self-assessment questionnaire completed by the students before and after the course. This questionnaire was designed to determine, for each element covered in the course, the student's assessment of its importance to him professionally and of his present understanding or skill in the area. Exhibits 15 and 16 show the instructions presented to students and a sample page from the questionnaire. With a self-assessment instrument of this type, used before and after the course, one can determine (a) how far the course has changed the attitudes of participants toward the importance of various course elements and (b) how far the student himself feels he progressed in his knowledge of various subject areas. The authors of this report are careful to point out that, when so many approaches are taken to the evaluation of a single course, there is a very real danger that students will find the evaluation "oppressive" and will rebel against it.

Exhibit 15

Instructions to Participants in a
Course on "Position Classification"
Organized by the U.S. Civil Service Commission/71

SELF-ASSESSMENT QUESTIONNAIRE

This is a self-assessment questionnaire and we will ask you to complete this before and after the course. The elements listed will be covered in the course. We would like your reaction to the importance of each element for successful performance of your job and what you feel to be your present level of understanding or skill on each element.

Please indicate beside each element the importance of the element for successful performance of your job by marking in the appropriate box:

- \- if that element is of no importance or if only minor importance
- o if that element is of some importance but not essential
- \+ if that element is essential to effective performance

In addition, please circle a number value on the scale beside each element which best indicates the understanding or skill you feel you presently possess. You may circle any number on the scale. The scale is as follows:

No understanding or skill			Moderate understanding or skill					Completely adequate	
0	1	2	3	4	5	6	7	8	9	10

Understanding is evidenced by the ability to define, describe or cite the major aspects of the element.

Skill is evidenced by the ability to apply the major aspect of the element.

In addition to the questionnaire, please answer the following:

Do you presently have classification responsibilities /__/ Yes /__/ No

Understanding of the following fundamentals of employee development:

1. Terminology	/__/	0 1 2 3 4 5 6 7 8 9 10
2. Basic Legislation	/__/	0 1 2 3 4 5 6 7 8 9 10
3. Structure of the System	/__/	0 1 2 3 4 5 6 7 8 9 10
4. Methodology, Techniques and Procedures	/__/	0 1 2 3 4 5 6 7 8 9 10
5. General Policy and Philosophy of the System	/__/	0 1 2 3 4 5 6 7 8 9 10
6. Trends in the Field	/__/	0 1 2 3 4 5 6 7 8 9 10
7. Principles	/__/	0 1 2 3 4 5 6 7 8 9 10

Skill in the following:

48

8. The Desk Audit	/__/	0 1 2 3 4 5 6 7 8 9 10
9. Analysis of Position Classification	/__/	0 1 2 3 4 5 6 7 8 9 10
10. Analysis of Job Information	/__/	0 1 2 3 4 5 6 7 8 9 10
11. Evaluation of Positions	/__/	0 1 2 3 4 5 6 7 8 9 10
12. Writing of Position Descriptions	/__/	0 1 2 3 4 5 6 7 8 9 10
13. Writing of Evaluation Statements	/__/	0 1 2 3 4 5 6 7 8 9 10
14. Dealing Effectively with Line Officials and Employees	/__/	0 1 2 3 4 5 6 7 8 9 10

Exhibit 16

Sample Page From Student Self-Assessment Questionnaire Used in a Course by the U.S. Civil Service Commission[71]

5. DESIGN OF EVALUATION INSTRUMENTS

It has been pointed out by a number of writers that the evaluation of educational programmes is not "research", at least in the strict sense of that term. In fact, as Stufflebeam[68] has indicated, there is considerable controversy as to whether experimental design has any real relevance to educational evaluation. He points out that two polar positions exist: the belief that experimental design has no relevance to the evaluation of education and the belief that educational evaluations are valid only if they are based on careful experimental design. Stufflebeam himself believes that experimental design has only a very limited application in the evaluation of education. While a knowledge of research methods and principles of experimental design do have value in planning and executing even a simple evaluation programme, these subjects are outside the scope of the present report. A useful account of experimental design and the statistical analysis of survey data, as applied to the field of education, is given in a recent book by Isaac and Michael[28], another, somewhat less detailed in Dizney[13], and an excellent description of the design of educational tests in terms of their statistical validity and reliability is given in Evaluation Guidelines[15], a publication of the Nevada State Department of Education. The subject is treated in numerous other books on the evaluation of educational programmes.

Of particular relevance to the type of evaluation discussed in this report is the design of effective questionnaires. Substantial volumes have been written on the subject of questionnaire design. Books by Oppenheim[47] and by Payne[49] are recommended, as is a concise summary by Whitney and Brandenburg[73]. Although we cannot go into great detail here on the subject of questionnaire design, it certainly seems desirable to mention some general principles.

It is possible to distinguish two major types of questions that may be used in an evaluation instrument: the open-ended question and the closed-ended question. In the former, the respondent answers freely in narrative style while, in the latter, an answer is selected from a list of alternatives supplied by the investigator. The division is not an absolute dichotomy, however, for some closed-ended questions may require an open-ended explanation. In this case the respondent is asked for more than his opinion. For example, if he is critical of the use of audiovisual materials in a course, he may be asked to explain the reasons for his criticism and/or suggest ways in which these materials might have been used more effectively. Early in this report, the importance of microevaluation was stressed. Microevaluation is an analytical or diagnostic level of evaluation in which one tries to discover not only what is happening but why it is happening and what might be done to improve a situation that may not now be satisfactory. It would be useful to discover that most students participating in a particular course were dissatisfied with the way that audiovisual materials were used. But this in itself only indicates a specific area of dissatisfaction. It does not precisely identify the reason for the dissatisfaction. One needs to know the underlying causes of dissatisfaction in order to identify an appropriate corrective action. For this reason, one may want to make some use of open-ended questions or at least open-ended "elaborations" on certain closed responses. Open-ended questions have a number of other advantages, including the fact that they can produce a wide variety of responses, some of which may be completely unanticipated by the investigator, that they are largely free from the influence of the investigator (it is difficult to deliberately bias the responses to an open-ended question in a printed questionnaire), and that they give respondents a chance to express their own opinions in any way they want.

The disadvantages of open-ended questions are also the advantages of questions of the closed-ended type. The latter produce greater uniformity of response, provide results that are easier to tabulate and perhaps to interpret, reduce or eliminate problems of terminology and definition, and, possibly most important of all, reduce the amount of effort that the respondent must put into completion of the questionnaire. In a largely closed-ended questionnaire one can ask more questions and perhaps expect a higher rate of response. In questionnaires for course evaluation it will probably be necessary to use a mixture of closed-ended and open-ended questions. It will be desirable to use questions of the closed-ended type wherever possible, restricting the open-ended questions to situations in which it is important to get the respondent's own interpretation or his suggestions on how a situation might be improved.

Questionnaires are designed to gather two types of data: factual and opinions or attitudes. Both types might be collected in the evaluation of a course or workshop. The evaluation of learning may primarily involve factual questions and factual responses. However, in other aspects of the formative and summative evaluation of a course, the main emphasis may be on the attitudes of students. It is important, therefore, that we give some consideration to the subject of attitude scaling.

The simplest scale, clearly, is a dichotomous one, a scale in which the respondent simply chooses between two alternatives (e.g., YES/NO or AGREE/DISAGREE). The dichotomous scale gives the respondent least flexibility in the replies. It is suitable only for questions in which, for the majority of respondents at least, the answer will be clear cut. Frequently a question of this type can be used to lead in to other questions. That is, a yes/no or agree/disagree question may be used as a screening question. For example:

Did you attend any of the sessions on microcomputers? NO /____/

YES /____/

A respondent answering affirmatively is then asked further questions about participation in these sessions. If he answers negatively he is led on to a different section of the questionnaire.

The dichotomous scale should not be used if it seems likely that many respondents may want to qualify their responses ("Yes, but..." or "Yes, except..."). In this situation it is better to go to some more flexible response scale.

Multiple choice questions fall between the two extremes of the completely open-ended questions and the dichotomous question. The multiple choice question is particularly suitable for use in determining attitudes or opinions because it provides for gradations in these attitudes. There are several possible scales that can be used with multiple choice questions.

Multiple choice questions involve the use of checklists from which the respondent makes a choice. Three broad types of checklist can be identified as follows:

1. the list from which the respondent chooses the single best or most appropriate response;

2. the list from which the respondent chooses all responses that apply to his particular situation;

3. the list that the respondent is asked to put in order of preference; that is, to produce a ranking of responses.

Whichever of these is used, it will sometimes be necessary to include an "other" category to take care of situations that were not anticipated in advance by the designer of the questionnaire.

It is in the arrangement of multiple choice checklists that a number of possible scales are available. The simplest is one in which the respondent selects a single answer from a list of possibilities as in the following example:

Select the statement that best indicates the novelty of the material presented in this course.

Most of the material was new to me ☐

About half was new to me ☐

Less than half was new to me ☐

None of the material was new to me ☐

Under certain circumstances it may be possible to organize the checklist in the form of a grid or matrix. This is particularly appropriate in a situation in which the same set of responses applies to several different situations, as in the following example:

For each of the instructional methods used in the course, please check the box that best describes your reaction to its value as a learning experience

	Great Value	Some Value	Little Value	No Value
Transparencies				
Films				
Demonstrations				
Practical exercises				
Discussion periods				

A popular form of attitude scale is the so-called "Likert scale", which is quite appropriate to many applications in the evaluation of courses or workshops. In this type of scale the respondent indicates his degree of agreement with some positive or negative statement. The scale frequently extends from "Strongly Agree" to "Strongly Disagree". It may consist of only four categories (Strongly Agree, Agree, Disagree, and Strongly Disagree) or it may include a fifth category, Undecided, that appears at the centre of the scale. Omission of the Undecided category forces a choice among the four positive/negative responses. In some cases the Likert scale may be expanded to seven response categories. It can also be used with Like-Dislike, Accept-Reject or other response ranges.

The Likert scale may be applied to a complete set of statements about a course or about an instructor. About half of these statements should be favourable and about half unfavourable. The scoring of the negative statements is the reverse of the scoring of the positive statements. Assuming a four-point scale, one could score the positive and negative statements according to the following scheme:

	SA	A	D	SD
Positive statement	4	3	2	1
Negative statement	1	2	3	4

In this scoring method the most positive attitude is represented by the highest score. When all the scores for all responses are averaged, those scores that are closest to four are the ones that represent the most positive attitude.

Exhibits 17 and 18 show examples from a Course Evaluation Questionnaire (CEQ) used by the University of Illinois at Urbana-Champaign. The form incorporates both open-ended and closed-ended questions. There are up to 26 closed-ended questions, some involving positive statements and some negative, all presented on the Likert scale. Note how the same scale is used for different ranges of responses: agree/disagree, excellent/very poor, and so on. The closed-ended responses are all coded for easy data reduction by computer. The data collected in the form of responses are manipulated by statistical programmes that calculate means, standard deviations and other values. Exhibit 19 presents a sample of the data collected in the evaluation of a particular course, this based on student responses to eight questions. The responses to each question are tabulated and the data are then grouped into broad categories (attitude of instructor, method of instruction, and so on). Finally, a total score for the course is derived. The maximum possible score for any question (and for any category) is 5.00. Data are presented as both totals and percentages. Thus, for question 4, 23 students (68% of respondents) strongly agreed that the instructor was conscientious. The instructor scored a 4.6 on this question.

A very similar scale is one that has implied numerical values; for example, a five point scale from "high" to "low" as illustrated in Exhibit 20. An alternative version (Exhibit 21) is the so-called "Smiley chart". In this case, the respondents choose the facial expression that best represents their feelings.

More complex is the Thurstone scale. To implement this, one must construct a series of statements about the facet of the course for which student attitudes are solicited. Some of these statements should be favourable to the course, some unfavourable, and some neutral. The Thurstone scale needs to be "calibrated" before it can be applied in the actual survey. This may be achieved by having the set of statements pretested by a group of individuals who are "similar" to those who will participate in the actual evaluation. Each of these "judges" ranks the statements into, say, ten levels according to the degree to which the statement is considered positive or negative. For each of the statements, the distribution of the scaled values assigned by the various judges is plotted. The median of this distribution for each statement is then located. The more "ambiguous" statements (as indicated by major discrepancies among the judges as to where the statement appears on the favourable-unfavourable scale) are eliminated. From the remaining statements, a set (perhaps about twenty) is selected such that it contains statements that range all over the ten-point scale. The scale value for each statement selected is the mean of the rankings of the judges. The statements are put in

Exhibit 17

Closed Questions from a Course Evaluation Questionnaire Used at the University of Illinois

Reproduced by permission of the Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign. Copyright, University of Illinois, Board of Trustees, 1977.

1. RATE THE COURSE CONTENT	EXCELLENT	○○○○○	VERY POOR
2. RATE THE INSTRUCTOR	EXCELLENT	○○○○○	VERY POOR
3. RATE THE COURSE IN GENERAL	EXCELLENT	○○○○○	VERY POOR
4. WAS THE PROGRESSION OF THE COURSE LOGICAL AND COHERENT FROM BEGINNING TO END?	YES, ALWAYS	①②③④⑤	NO, SELDOM
5. DID THIS COURSE DUPLICATE OTHER COURSES IN THIS AREA?	CONSIDERABLY	①②③④⑤	NOT AT ALL
6. HOW WELL DID THE INSTRUCTOR COORDINATE DIFFERENT ACTIVITIES OF THIS COURSE?	VERY WELL	①②③④⑤	RATHER POORLY
7. THIS COURSE WAS CREATIVELY PLANNED.	STRONGLY AGREE	①②③④⑤	STRONGLY DISAGREE
8. HOW WELL DID THE READING, LECTURES, AND DISCUSSIONS COVER ANNOUNCED OBJECTIVES?	BALANCED COVERAGE	①②③④⑤	UNBALANCED COVERAGE
9. THE COURSE CONTENT WAS:	TOO THEORETICAL	①②③④⑤	TOO APPLIED
10. DID THE INSTRUCTOR PRESENT MATERIAL THAT WAS NOT COVERED IN OUTSIDE READINGS?	YES, OFTEN	①②③④⑤	NO, SELDOM
11. THE INSTRUCTOR CHANGED APPROACHES WHEN THE OCCASION DEMANDED IT.	STRONGLY AGREE	①②③④⑤	STRONGLY DISAGREE
12. I NEEDED MORE DIRECTION.	STRONGLY AGREE	①②③④⑤	STRONGLY DISAGREE
13. THE INSTRUCTOR DEFINED THE OBJECTIVES OF DISCUSSION.	ALMOST ALWAYS OCCURED	①②③④⑤	ALMOST NEVER OCCURRED
14. HOW OFTEN WERE YOU REQUIRED TO DO WHAT THE INSTRUCTOR WAS "INTO"?	ALMOST ALWAYS	①②③④⑤	ALMOST NEVER
15. RATE THE TEXT(S) USED IN THIS COURSE.	EXCELLENT	①②③④⑤	POOR
16. WERE READINGS WELL SELECTED?	YES, ALL VERY GOOD	①②③④⑤	NO, ALL VERY POOR
17. WERE READING ASSIGNMENTS RELEVANT TO CLASS PRESENTATIONS?	YES, ALWAYS	①②③④⑤	NO, ALMOST NEVER
18. ADEQUATE TIME WAS PROVIDED FOR COMPLETING ASSIGNMENTS.	ALWAYS	①②③④⑤	SELDOM
19. COMPLETING WRITTEN ASSIGNMENTS WAS A GOOD USE OF MY TIME AND EFFORT.	STRONGLY AGREE	①②③④⑤	STRONGLY DISAGREE
20. WERE WRITTEN ASSIGNMENTS GRADED FAIRLY?	YES, QUITE FAIR	①②③④⑤	NO, VERY UNFAIR
21. HOW BENEFICIAL WERE THE HOMEWORK ASSIGNMENTS?	VERY BENEFICIAL	①②③④⑤	JUST BUSY WORK
22. DID INSTRUCTIONAL MATERIALS APPEAR TO BE CONSCIENTIOUSLY PREPARED OR CHOSEN?	YES, ALWAYS	①②③④⑤	NO, SELDOM
23. AUDIO-VISUAL PROCEDURES WERE LOGICALLY INTEGRATED WITH THE REST OF THE COURSE.	STRONGLY AGREE	①②③④⑤	STRONGLY DISAGREE
24. HOW EFFECTIVE WAS THE INSTRUCTOR IN PREPARING STUDENTS FOR EXAMS?	VERY EFFECTIVE	①②③④⑤	VERY INEFFECTIVE
25. DID THE INSTRUCTOR HAVE A REALISTIC DEFINITION OF EXCELLENT PERFORMANCE?	YES, VERY REALISTIC	①②③④⑤	NO, VERY UNREALISTIC
26. WERE REQUESTS FOR RE-GRADING OR REVIEW HANDLED FAIRLY?	YES, ALMOST ALWAYS	①②③④⑤	NO, ALMOST NEVER

Exhibit 18

Open-ended Questions from a Course Evaluation Questionnaire Used at the University of Illinois

Reproduced by permission of the Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign. Copyright, University of Illinois, Board of Trustees, 1977.

A What are the major strengths and weaknesses of the instructor?	
B What aspects of this course were most beneficial to you?	
C What do you suggest to improve this course?	
D Comment on the grading procedures and exams.	

DEPARTMENTAL CORE ITEM RESULTS	RESPONSE LABELS, PERCENTS & FREQUENCIES						SUMMARY STATISTICS		RELATIVE COMPARISON OF CLASS MEAN
							MEAN MDN	SD CONS	ECONOMICS FACULTY
CATALOG ITEM 278									!LOW ! ! HI!
4 THE INSTRUCTOR WAS CONSCIENTIOUS ABOUT HIS/HER INSTRUCTIONAL RESPONSIBILITIES.	STRONGLY AGREE	--	--	--	STRONGLY DISAGREE	OMIT			LOW! AVG! AVG !AVG!HI
WEIGHTING SCHEME 54321	68	26	6	0	0	0	4.6	0.59	
	(23)	(9)	(2)	(0)	(0)	(0)	4.6	AVG	OOXOOO
CATALOG ITEM 101							MEAN	SD	
5 THE GRADING PROCEDURES FOR THE COURSE WERE:	VERY FAIR	--	--	--	VERY UNFAIR	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	12	29	29	21	9	0	3.1	1.14	LOW! AVG! AVG !AVG!HI
	(4)	(10)	(10)	(7)	(3)	(0)	3.2	LOW	XOO
CATALOG ITEM 103							MEAN	SD	
6 HOW WELL DID EXAMINATION QUESTIONS REFLECT CONTENT AND EMPHASIS OF THE COURSE?	WELL RELATED	--	--	--	POORLY RELATED	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	15	41	29	9	6	0	3.5	1.03	LOW! AVG! AVG !AVG!HI
	(5)	(14)	(10)	(3)	(2)	(0)	3.6	AVG	OOXOOO
CATALOG ITEM 5							MEAN	SD	
7 WAS THE PROGRESSION OF THE COURSE LOGICAL AND COHERENT FROM BEGINNING TO END?	YES, ALWAYS	--	--	--	NO, SELDOM	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	38	50	9	3	0	0	4.2	0.73	LOW! AVG! AVG !AVG!HI
	(13)	(17)	(3)	(1)	(0)	(0)	4.3	AVG	OOXOO
CATALOG ITEM 244							MEAN	SD	
8 HOW WOULD YOU CHARACTERIZE THE INSTRUCTOR'S ABILITY TO EXPLAIN?	EXCELLENT	--	--	--	VERY POOR	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	53	35	3	9	0	0	4.3	0.89	LOW! AVG! AVG !AVG!HI
	(18)	(12)	(1)	(3)	(0)	(0)	4.5	AVG	OOXOO
CATALOG ITEM 325							MEAN	SD	
9 THE INSTRUCTOR MOTIVATED ME TO DO MY BEST WORK.	ALMOST ALWAYS	--	--	--	ALMOST NEVER	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	35	26	26	12	0	0	3.9	1.03	LOW! AVG! AVG !AVG!HI
	(12)	(9)	(9)	(4)	(0)	(0)	3.9	AVG	OOXO
CATALOG ITEM 191							MEAN	SD	
10 DID THIS COURSE INCREASE YOUR INTEREST IN THE SUBJECT MATTER?	YES, GREATLY	--	--	--	NO, NOT MUCH	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	44	50	3	3	0	0	4.4	0.68	LOW! AVG! AVG !AVG!HI
	(15)	(17)	(1)	(1)	(0)	(0)	4.4	AVG	OOXOO
CATALOG ITEM 169							MEAN	SD	
11 DID THIS COURSE IMPROVE YOUR UNDERSTANDING OF CONCEPTS AND PRINCIPLES IN THIS FIELD?	YES, SIGNIFICANTLY	--	--	--	NO, NOT MUCH	OMIT	MDN	CONS	!LOW ! ! HI!
WEIGHTING SCHEME 54321	47	44	9	0	0	0	4.4	0.64	LOW! AVG! AVG !AVG!HI
	(16)	(15)	(3)	(0)	(0)	(0)	4.4	AVG	OOOXOO

P1188-00

Exhibit 19

Sample of Data from a Course Evaluation Questionnaire

Reproduced by permission of the Measurement and Research Division, Office of Instructional Resources, University of Illinois at Urbana-Champaign. Copyright, University of Illinois, Board of Trustees, 1977.

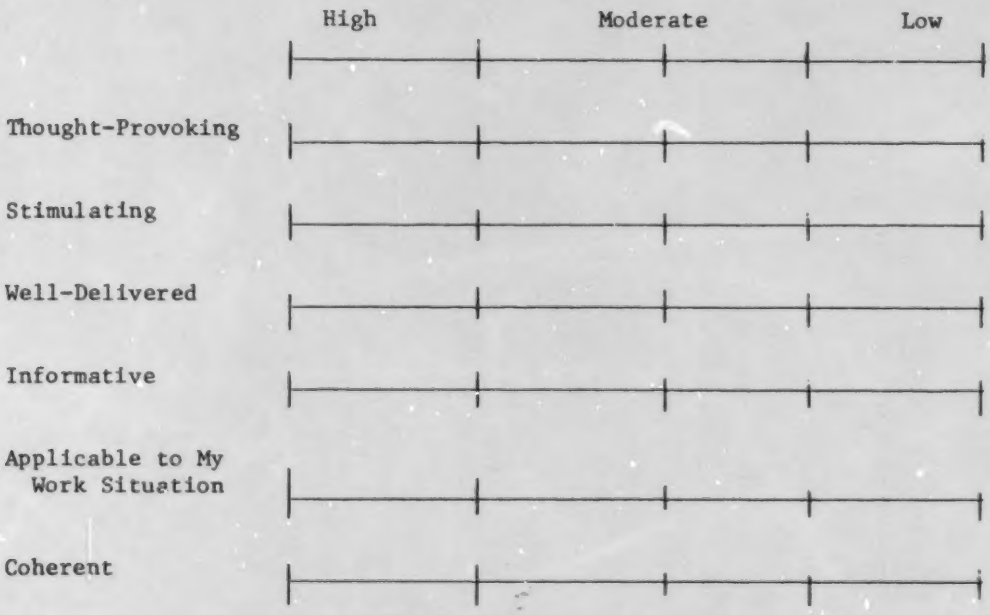
Exhibit 20

Typical Numerical Scale Used in an Evaluation Instrument

Reproduced from Greenfest and Bragg[21] by permission of the Gerontology Center, Pennsylvania State University

Speakers

I found _____'s presentations



Overall, this speaker related to me:

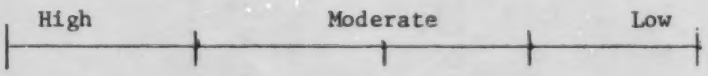


Exhibit 21

Use of the Smiley Chart

Reproduced from Slater[58] by permission of Aslib

Various aspects of the course are covered in the chart below. Please fill it in by putting a tick, for each aspect or question, under the face that most resembles your own reaction. You can add any explanatory comments in the column on the right of the chart – we would be very pleased to have them.

	QUESTION/ASPECT OF COURSE						YOUR COMMENTS
(a)	Your general opinion of the courses as a whole?						
(b)	Did the course achieve its aims, as set out in course brochure?						
(c)	To what extent were your specific aims met?						
(d)	How do you feel about the length of the course?						too long too short – – –
(e)	Were you happy about the main subjects selected as lectures?						
(f)	Did you feel that these subjects were covered in the right way?						
(g)	How do you feel about the intellectual level of the course?						too low – – – too high – – –
(h)	How did you react to the visual aids used?						no answer, as none used – – –
(i)	How do you feel about facilities and arrangements for attenders?						

random sequence on the questionnaire and the respondent simply checks off those he agrees with. A respondent's score is the sum of the scale values for the statements checked, the higher the score the more positive the response. The Thurstone scale is useful in determining the attitudes of individual respondents and perhaps to determine whether the attitudes have changed as a result of some additional experience. One could, for example, use the Thurstone scale to measure some attitude among students before and after taking a particular course.

A third type of scale is the so-called semantic differential. This scale involves the use of bipolar adjectives (e.g., weak/strong, good/bad, important/unimportant). The respondent is asked to rate a particular activity by marking a point on a scale between the two extremes represented by the bipolar adjectives, as in the following:

good _____ _____ _____ _____ _____ _____ _____ bad

A scale on which a respondent marks some point is frequently referred to as a "graphic scale". A whole group of such graphic scales is created in order to measure attitude towards a particular event or activity. The semantic differential is readily converted to a numerical score reflecting the degree of "positiveness" in the attitude of the respondent, as in the examples:

weak	1	2	3	4	5	6	7	strong
complete	7	6	5	4	3	2	1	incomplete

The attitude of a student towards a course or an instructor can be evaluated by a semantic differential incorporating a whole series of graphic scales of this type (see, for example, Carmichael et al.[17]). An example is given in exhibit 22. Any type of scale in which the respondent is required to indicate degree of satisfaction or approval is likely to be quite applicable to questionnaires for the evaluation of educational offerings.

In all of the scales so far discussed, the respondent checks off the single most appropriate response or marks a particular point on a scale of values. It is also possible to include questions in which the respondent must select all statements he agrees with. A scale widely used in the evaluation of courses is the Kropp-Verner scale[36], which consists of a series of eighteen statements, some positive and some negative. The respondent checks all the statements he agrees with. Use of the Kropp-Verner scale is illustrated in Exhibit 23 and in Question 9 of the survey instrument given in Appendix 5A. The results of the application of this scale are tabulated in the report appearing in Appendix 7A.

A final scale worth mentioning is the ranking scale. In this, a list of alternatives is given and the respondent is asked to rank these in order of priority. Again, rankings of this type have possible utility in the evaluation of courses or workshops. For example, one could ask a student to rank the topics covered in order of their relevance to him, or to rank, in order of effectiveness, several approaches to instruction used in a course.

One approach to ranking is to have the respondent put a list of items (e.g., all the teaching methods used in a course) in numerical order to reflect his perception of importance. A second is to have the respondent simply group the items into a few broad classes, such as Important, Less Important, Unimportant.

Exhibit 22

Use of the Semantic Differential for Judging Teacher Attitudes Toward Consumer Education

Reproduced from A. D. Grotelueschen et al. Evaluation in Adult Basic Education.[24] Copyright 1976, Interstate Printers and Publishers Inc. Reproduced by permission of the publisher.

CONSUMER EDUCATION

Worthless	1	2	3	4	5	6	7	Valuable
Important	1	2	3	4	5	6	7	Unimportant
Useless	1	2	3	4	5	6	7	Useful
Easy	1	2	3	4	5	6	7	Difficult
Theoretical	1	2	3	4	5	6	7	Practical
Inexpensive	1	2	3	4	5	6	7	Expensive
Concrete	1	2	3	4	5	6	7	Abstract
Uninteresting	1	2	3	4	5	6	7	Interesting
Meaningless	1	2	3	4	5	6	7	Meaningful
Exciting	1	2	3	4	5	6	7	Boring
Relevant	1	2	3	4	5	6	7	Irrelevant

Exhibit 23

The Kropp-Verner Scale

Reproduced from Kropp and Verner[36] by permission of the American Association for Adult and Continuing Education

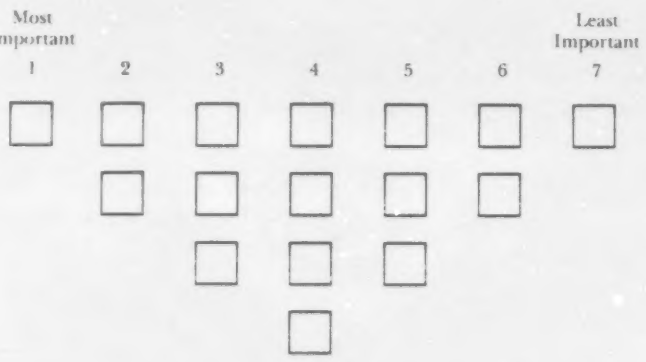
Please follow directions carefully: Read all twenty of the following statements. Check as many statements as necessary to describe your reaction to the Institute.

1. _____ It was one of the most rewarding experiences I have ever had.
2. _____ Exactly what I wanted.
3. _____ I hope we can have another one in the near future.
4. _____ It provided the kind of experience that I can apply to my own situation
5. _____ It helped me personally.
6. _____ It solved some problems for me.
7. _____ I think it served its purpose.
8. _____ It had some merits.
9. _____ It was fair.
10. _____ It was neither very good nor very poor.
11. _____ I was mildly disappointed.
12. _____ It was not exactly what I needed.
13. _____ It was too general.
14. _____ I am not taking any new ideas away.
15. _____ It didn't hold my interest.
16. _____ It was much too superficial.
17. _____ I leave dissatisfied.
18. _____ It was very poorly planned.
19. _____ I didn't learn a thing.
20. _____ It was a complete waste of time.

(If you wish, add any comments on reverse side of this page.)

Strict numerical ranking has some problems associated with it. In general, only the values at the ends of the scale can be considered meaningful. People tend to be clear about what they consider useful or useless, but the middle ground contains a much less clearly defined range of values. Ranking of this kind gives a spurious appearance of quantification, because the items being ranked seldom can be considered strictly as elements from a homogeneous continuum. As Grotelueschen et al.[24] have pointed out, an item ranked 1 on a list may be only marginally more important, in the eyes of the respondent, than one marked 2, while the item ranked 3 is substantially less important than that ranked 2. Such differences, which may be significant, are just not revealed by numerical ranking procedures.

Grotelueschen et al. suggest that a ranking exercise tends to produce better results when the respondents are asked to manipulate cards or pieces of paper, as in the so-called Q-Sort technique. Given a set of cards, each one bearing an item to be ranked, respondents are required to place these cards on a numerical scale, as in the following example:



Note that several items can be given the same rank. Of course, once the cards have been arranged in this way, the results can be transferred to paper. It is reported that respondents seem to give more thought and concentration to a ranking task when presented in this more concrete form, re-arranging the items until they are finally satisfied with the results.

There is no particular reason why a questionnaire should use only one scale. It is perfectly acceptable to design an evaluation instrument in which a number of different scales, each most appropriate to a certain type of question, have been incorporated.

The discussion so far has largely been restricted to the subject of attitude scales. Such scales are valuable to determine the attitudes or opinions of students. They could also be used to assess the attitudes of students towards various facets of the subject matter of the course. Thus, if the course is intended to influence the students' attitudes in some way, we can test for attitude before and after the course. The construction of attitude scales is discussed and illustrated in books by Edwards[14] and by Shaw and Wright[55].

Another subject that deserves some attention here is the design of tests to measure the learning gained by the students in a course. One might be interested

in this type of evaluation for courses, not to grade students but to determine how effective the educational experience has been. The goal here should be to design reasonably objective tests of the learning acquired by students.

Gronlund[22] distinguishes the following major categories of tests of learning:

A. Test in which the student supplies the answer

   1. Essay

   2. Brief notes

   3. Fill-in-the-blanks

B. Tests in which the student selects the answer

   1. True-false

   2. Matching

   3. Multiple choice

The latter group, the selection-type tests, are more controlled and objective. The answer selected by the student is either correct or it is not. Grading of such tests is quite easy. The fill-in-the-blanks type of test also produces reasonably unequivocal responses. Essays are the most difficult form of test to grade objectively and "brief notes" (which may be regarded as "mini-essays") are almost as bad.

Use should be made of selection-type tests wherever possible. Not only are these more objective, and thus easier to grade, but selection-type tests require less time and thus one can afford to test more extensively over the full range of subject matter. Moreover, selection-type tests have another great advantage in the international environment. The student whose native tongue is not that of the course will be at less of a disadvantage in this type of test. In fact, it should be relatively easy to produce such a test in several languages to reduce this disadvantage even further.

With a little ingenuity most tests of learning can be reduced to the selection format. For example, consider the question:

> With what developments in information retrieval do you associate the following names:
>
> William Batten
> Calvin Mooers
> Mortimer Taube
> Eugene Garfield
> Hans Peter Luhn

This question could have been handled in many other ways. It is "natural" for example for use in a matching test, as follows:

Match the names in Column A with the developments in information retrieval shown in Column B. Put the appropriate number against each name.

A	B
William Batten	1. SDI
Calvin Mooers	2. Uniterm
Mortimer Taube	3. Citation indexing
Eugene Garfield	4. Peek-a-boo
Hans Peter Luhn	5. Notched cards

The question could also be handled, although somewhat more laboriously, in a true/false format:

Mortimer Taube was the first person to use peek-a-boo cards for information retrieval.

TRUE /___/

FALSE /___/

or in a multiple-choice approach:

Mortimer Taube:

a. invented citation indexing /___/

b. developed the first SDI system /___/

c. produced the Uniterm system /___/

d. invented the peek-a-boo principle /___/

e. pioneered the use of notched cards for information retrieval /___/

An alternative would be to have the student select all the statements that are correct in relation to some situation or phenomenon. For example:

Select all statements that you consider represent advantages that online database searching offer over offline database searching.

The search results can be printed out /___/

The searcher can interact directly with the database /___/

Microcomputers can be used /___/

A search can be performed for current awareness purposes /___/

Virtually immediate results can be achieved /___/

Requester and searcher can work together as the search is performed /___/

The search can be performed heuristically /___/

Only information specialists can perform such searches /___/

Most factual-type questions can be presented in closed-ended form according to one of these models.

Before any type of test of learning is designed one must obviously define the learning outcomes to be measured in the test. In the above discussion, a course or workshop whose prime purpose is the transfer of specific subject knowledge or skills has been assumed. For an educational endeavour of this kind, it should be possible to test objectively the extent to which the subject knowledge or skills are transferred. But a course or workshop may have other objectives. It may be designed, for example, to improve the problem solving or decision making skills of the participants. It is not inconceivable that a test of problem solving or decision making skills could be reduced to a selection-type exercise. It is more likely, however, that one will need to design something more elaborate. Tests of decision making or problem solving can be designed for many fields (e.g., problems involved in managing some operation) along the lines of the tests of decision making in health care that were mentioned earlier in this report.

A technique that might be used in the evaluation of courses designed to produce attitudinal changes, or to improve problem solving and decision making skills, is the student self-evaluation. This involves a self-rating scale. Students judge their own position on this scale before and after the course or workshop. Burstiner[5], for example, used a scale of 20 "creativity" traits and 20 "leadership" traits in evaluating a workshop on creative leadership. Each participant rated himself on each of these traits on a seven-point scale ranging from "extremely low" to "extremely high". Significant gains were recorded in both creativity and leadership scales after the course was completed. An example of a self-assessment questionnaire is given in Exhibits 15 and 16.

More complete discussions of the design of tests of learning are given in books by Gronlund[22, 23], by Ahmann and Glock[1], Bloom et al[2], and Lindquist[39].

6. EVALUATING SPECIFIC EDUCATIONAL ACTIVITIES

The earlier sections of this report have discussed, in a general way, the subject of evaluation of education and have looked at various possible approaches and methods. This section will be concerned more specifically with the evaluation of various activities (courses, workshops, seminars) and will present recommendations and guidelines on how these activities might be evaluated. It is assumed from now on that a decision has been made that a particular educational activity will take place and that this activity is relevant to the organizer's overall objectives in the area of education and training. We are not, therefore, concerned with evaluating the need for a particular course but only with evaluating the course, in a formative and summative way, as an instructional unit and as a component in a larger programme of educational activities.

As identified by Thiede[/69], five steps are involved in the evaluation of a particular activity:

(a) deciding what to evaluate,
(b) defining the behaviour desired,
(c) determining acceptable evidence,
(d) collecting evidence, and
(e) summarizing and evaluating the evidence.

These steps should be kept in mind throughout the planning of evaluation activities. As emphasized on several earlier occasions, it is important that consideration be given to evaluation from the very beginning of a programme. The staff of the planning body, having decided on some type of educational offering in a specific subject area, must proceed to (a) select an appropriate place and time for the meeting, and (b) select the instructor who is to have major responsibility for planning and implementing the programme. Once this is done, it will be the joint responsibility of the planning staff and the instructor to define clearly what is to be covered (i.e., the content), who is to participate (i.e., what type of person should be attending), what the general format of the programme should be (a lecture course, a workshop, or a combination of these), and its duration. At this point the instructor must prepare an outline of the subject matter to be covered and an indication of what level the material is to be presented at. At the same time, the organizers, working with the instructor, must draw up a list of requirements for participants in the programme (e.g., language facility, educational background, previous professional experience) and must decide how many participants will be accepted. Participants must be selected on the basis of their previous knowledge of the subject matter in relation to the level set for the course. It is as important to avoid overqualified participants as it is to avoid the underqualified. Some demonstration of abilities in the language in which the course will be given seems highly desirable. The general objectives of the course must also be clearly defined at this point. That is, a statement of desired outcomes will be prepared. Anticipated outcomes must clearly be related to both the proposed content and to the type of student who will attend the programme. To take one example, a Unesco course held in Katowice in 1974 can be depicted according to the following categorization:

Subject	Information retrieval and information retrieval systems.
Level	Introductory to intermediate.

Students	Proficiency in written and spoken English. Previous experience or training in scientific documentation. Responsibility (present or future) for training information workers in their own countries.
Major objective (outcome)	To present a state-of-the-art survey of modern methods of information storage and retrieval, and how this content can best be presented, so that participants will be equipped to plan or conduct similar courses in their own countries.

It will be important at this point for the instructor, together with the course organizers, to evaluate the general objectives set for the course (in relation to the other specified variables of type of instruction, type of student, location and duration of the course) to determine if they are, in fact, reasonable and realistic. The course at Katowice presented a good example of the need to evaluate stated objectives in relation to the other specified variables. Students selected to participate in this course had quite varied backgrounds. A few already had a reasonable knowledge of information storage and retrieval but the majority had little or no knowledge of the subject matter. For a completely introductory course on the subject, some of the participants were overqualified. Without previous knowledge of the subject it was unrealistic to expect this one three-week course to teach participants enough to allow them to present a similar course in their own countries. The course would, however, give enough background to allow participants, if they read and studied further at home, to develop one or more courses in this area. It should certainly have given all participants sufficient knowledge of the field that they could plan a programme in this subject area (i.e., develop a curriculum) even though they may not know enough to teach all aspects of the subject themselves. On the other hand, for those with previous knowledge of the subject matter, it would be possible to use the course as a "refresher". These students, with their previous knowledge reinforced by this course, could reasonably be expected to teach the subject on return to their own countries. This particular situation can be summarized as follows:

	Student background	Emphasis of the course	Major objective or outcome
1.	No knowledge of subject matter	Impart broad knowledge of subject matter	Allow student to develop curricula in this area and to develop plans for training programmes
2.	Good knowledge of subject matter already	Refresher course on subject with emphasis on content and how it should best be presented	Allow student to develop and teach specific courses in this field

Note how the pre-existing background of the student influences the direction that the course must take and the objectives that can reasonably be achieved. In the course at Katowice, where students of quite varied backgrounds participated, it was necessary to compromise. The instructors found it necessary to assume no previous background and to present a broad overview of the subject matter. As a result, the course is likely to have achieved, for most participants at least, the first objective above but not the second.

This case illustrates clearly that, in the initial planning of a course, it will be important to examine the broad objectives carefully and to choose the complete instructional staff, the mode of instruction, the duration of the course and (perhaps)

its location, and, most importantly, the students in such a way that these broad objectives have the best possible chance of being achieved.

It is recommended that all educational offerings make some provision for both formative and summative evaluation. At the very least, the summative evaluation should be based on overall student reaction. If at all possible, it would be desirable to arrive at some standard end-of-course evaluation form, one that could be used intact (or with only minor modification) in all courses offered by a particular organization. This would allow meaningful comparisons to be made from one programme to another. In addition, if the course is designed to produce competence in a particular subject area (as opposed to dealing with "attitudinal" matters), it would be highly desirable to incorporate some method of evaluating what each student has learned. It would also be desirable to attempt to assess the long-term effects of the course; i.e., the "behavioural changes" it produces. This may require a follow-up questionnaire administered to all participants some months after the course is concluded.

It is recommended that the instructor who has prime responsibility for the development of a particular course should develop a detailed outline of what the course will cover. The content must be clearly related to the overall course objectives and to the "level" of the participants. Once this outline is developed, the instructor should develop, for each section of the course, one or more behavioural objectives. For each behavioural objective, the instructor must develop some demonstration of achievement. Some of the behavioural objectives set for the Katowice course and the accompanying demonstrations of achievement are put forward as an example (see Exhibit 24). The more limited the objectives of the course, and the more practical its nature, the easier it will be to develop a reasonably unequivocal demonstration of achievement. For example, if a course was developed on thesaurus construction it would have one major behavioural objective (to teach students how to construct a thesaurus according to ISO guidelines) and one major demonstration of achievement (the ability of the student to construct a sample thesaurus in a subject field related to his own interests). If the subject coverage is quite diffuse, however, it will be necessary to have a number of behavioural objectives and a number of demonstrations of achievement, as was the case with the Katowice course, which attempted a broad overview of information storage and retrieval. The less concrete or practical the subject matter of the course, the more difficult it will be to develop unequivocal demonstrations of achievement.

The specific behavioural objectives developed for a course will have an important role to play in its overall evaluation. If the instructor, and the organizers, consider it appropriate, these behavioural objectives can be used in the evaluation of learning acquired. That is, the students can be required to demonstrate achievement of each objective during the course itself. At the very least, however, the behavioural objectives, and stated demonstrations of achievement, should be available to all students so that, as each segment of the course is completed, they can ask themselves whether or not they feel that the behavioural objective has been satisfied and that they could in fact demonstrate its achievement appropriately. In the Katowice course the behavioural objectives were mainly used in this way although in some cases practical exercises were used as a demonstration of student achievement (e.g., exercises in indexing, the construction of searching strategies, and the development of a specimen thesaurus page).

If the behavioural objectives are not themselves to be used in the evaluation of student learning, it will be necessary to develop some other procedure for this. If the subject matter of the course is susceptible to this kind of treatment, a set

Exhibit 24

Some of the Behavioural Objectives Developed for a Course in Information Storage and Retrieval

Objectives	Demonstration of Achievement
1. Understand the input and output operations involved in modern computer-based IR systems	1. Be able to describe the functions of such a system in terms one of your own students would understand.
2. Recognize the advantages of on-line and off-line IR system.	2. Be able to enumerate these advantages and discuss them clearly and concisely.
3. Understand how search strategies are constructed and understand the Boolean logic involved in such strategies.	3. Be able to conceptually analyse a request for information, represent the relationships graphically by Venn diagrams, reduce the conceptual analysis to a set of index terms and logical operators, and reduce eventually to a simple equation.
4. Understand the distinctions between pre-coordinate and post-coordinate retrieval systems and the advantages and disadvantages of both types.	4. Be able to categorize any system as pre-coordinate or post-coordinate. Be able to write a paper clearly distinguishing the two types, with examples, and comparing the pros and cons of each.
5. Understand the major factors governing the performance of any IR system.	5. Be able to list these factors in a table, or reduce them to the form of a diagram, and be able to discuss them. For a particular system of given characteristics, be able to identify the factors most likely to be influencing its performance.
6. Understand the importance of machine-readable data bases, the range of available data bases, and the effect of the availability of these data bases on the provision of information services.	6. Be able to prepare a plan for the implementation of an information service, using existing data bases, in your own country.

of "before and after" questions can be devised. Despite the caveats presented earlier, a simple set of test questions used in Katowice before and after the course seemed reasonably successful in providing a relatively gross indication of the extent to which learning had occurred. There was little evidence there that the participants expected to be retested and, indeed, that the lecturers placed undue emphasis on the subject areas covered by the test questions. Indeed, the score on the end-of-course test in Katowice, although it showed a definite improvement, was far from perfect. It would have been better, however, if the lecturers had developed two sets of questions, one to be used before the course began and the second after it had been completed. From a complete set of questions, one half is drawn at random for the pre-test and the remainder is used in the post-test.

The pre-course questionnaire (see Exhibit 25 for an example) should be used to determine the motivations of the students and what they hope to achieve from their participation. The questions used to test subject knowledge can be incorporated into the pre-course questionnaire. They can be used not only to measure the learning acquired by students during the course but also to give the lecturers a better understanding of what the students know at the time they enter the programme. Some modifications of the course content may need to be made on the basis of the results of a pre-course assessment of this type.

The formative element is perhaps the most important in the entire evaluation process. To ensure the maximum level of success, the teaching staff must obtain feedback from students as the course is progressing. Although some such feedback will be obtained informally during the normal course of events, it is recommended that a formal evaluation be conducted periodically as the course progresses. For this purpose it will be necessary to devise and apply a "reaction questionnaire". Such a form cannot be completely standardized for all courses, because each will have somewhat different features, but the major elements could remain more or less the same from one course to the next. These elements were enumerated in Exhibit 4 and specimen questionnaires for formative evaluation purposes are illustrated in Appendix 1. The weekly reaction form used in Katowice (see Appendix 1) was quite useful to the lecturers in determining how the course was progressing and what might later be done to improve it.

An end-of-course evaluation questionnaire will be the major component in the terminal evaluation of the programme. Some of the more important questions to be answered in such a study are given in exhibit 7 and specimen questionnaires are shown in Appendix 5. The form used in Katowice was considered by the lecturers to be quite successful in measuring the overall reaction of participants to the course as a whole. The end-of-course evaluation questionnaire may also include the set of questions used to test the learning acquired by students.

As part of the evaluation it may be desirable to make some use of an independent observer. The observer will mix with students to obtain their reactions, observe the teaching methods used, and generally look for problems and ways in which the course might be improved. The independent observer should prepare his own report on the course. This report will supplement or complement the report prepared by the lecturers. It is obvious that the observer needs to be knowledgeable in the subject matter of the course. He must also be sociable, able to talk to participants in an easy way and to encourage them to speak freely. The observer should preferably attend the entire course. If this is impractical, in the case of a long course he or she should attend selected days throughout the entire course, including the first two days and the last two days. It is not enough to have the observer appear at the end of the course. Without knowing what went on earlier, he will not get a complete

Exhibit 25

Pre-Course Background Questionnaire

Reproduced from Stake et al[64] by permission of the University of Illinois

This checkup is intended to help project personnel plan the project to take account of your expectations and needs. If you think it asks the wrong questions or you want to add more comments, please feel free to use the back of the sheet to say so.

1. Name: _____

2. School: _____

3. Grade Levels Currently Taught: _____

4. Subject Specialization (if any): _____

5. Special Responsibilities (if any): _____

6. Number of Years of Teaching Experience: _____

7. What topics (if any) relating to the environment have you studied extensively? _____

_____

8. What special areas of expertise (if any) do you have in such topics?

_____

9. What kinds of experiences have you had related to environmental issues?

_____

_____

10. What kinds of environmental concepts (if any) have you recently taught?

_____

_____

11. What is your most important reason for participating in this project?

_____

_____

12. What special concepts and/or skills do you hope to learn from this project? _____

_____

picture, and the value of his contribution will be correspondingly reduced. Indeed, his interpretation and assessment of the course may be invalid if he has been unable to witness the entire programme as a logical sequence of events.

The final element in a complete evaluation will be an attempt to assess the impact of the course on those who attended, i.e., to determine its behavioural effects. This will involve at least a follow-up questionnaire mailed to all participants some months after the course is completed. Perhaps a full calendar year should elapse before such a questionnaire is administered. This questionnaire will be used to determine whether or not the students have made any use of the learning they acquired. Have the students done further work on their own in the subject matter of the course? Have they incorporated any of the material into their own teaching? Have they developed courses or workshops themselves in the subject area of the course? Has the course had any influence on their own planning of curricula or training programmes? If the course was directed to practitioners in the field, has it had any impact on the way they manage their own activities? Various possible approaches, including use of telephone interviews, were described in section 3.

It will probably be impossible to standardize such a questionnaire because each course or workshop will have somewhat different long-term objectives. Various examples were given in section 3 and the follow-up questionnaire developed for use with the participants at Katowice is shown in Appendix 6.

Of course, it is not enough to simply collect evaluative data. Some use must be made of it, hopefully to improve educational programmes in the future. For each course or workshop it will be the responsibility of the teaching staff to react to the formative elements in the evaluation. They, and the course itself, must be sufficiently adaptable that changes can be made, if needed, as the programme progresses. Once the course is completed, it will be the responsibility of the senior instructor to prepare a comprehensive report on the evaluation that has been conducted. This report must summarize the evaluation data clearly, preferably with as much as feasible in tabular form, it must interpret the results as far as possible (in particular, any that can be regarded as unexpected or anomalous), and it must include the lecturers' suggestions as to how, in retrospect, the course might have been improved and their recommendations on changes that should be made in future courses of this type. Examples of final reports are included as Appendix 7. The subject of reporting is treated further in section 8.

In summary, the complete evaluation of a course should involve the following elements:

1. Close co-operation between the senior instructor and the organizers in evaluating the objectives set for the course and in defining its major parameters (size and composition of the teaching staff, size and composition of the student body, course content, duration and location) in such a way that there is a reasonable expectation that the course objectives can be met.

2. The senior instructor should develop appropriate questionnaires, along the lines suggested in this report, for the ongoing formative evaluation of the course. Ongoing evaluation should take place in all educational offerings unless they are of such short duration that such an element seems impractical.

3. The senior instructor should consult with the organizers to determine the specific methods that will be used to assess the learning acquired during the course. The instructor should develop specific behavioural objectives for the course and for

each of these prepare one or more demonstrations of achievement. If considered appropriate by the organizers, and if the duration of the course permits it, these demonstrations of achievement can be used in a formal way throughout the course to assess the learning taking place. If this approach is not followed, the instructor should develop one or more sets of test questions that can be used, before and after the course, to get some indication of how much learning has actually occurred. Test questions of this kind (which must also relate to behavioural objectives) are likely to be quite effective for courses in which specific subject knowledge is to be imparted. For courses or workshops that are intended to produce "attitudinal" changes rather than to impart specific subject knowledge, the behavioural objectives and appropriate demonstrations of achievement may be used to judge how successful the course has been, or some test of attitudinal change (see section 2) can be employed.

4. Wherever possible, it would be desirable to have an independent observer present at the course. The respective roles of the teaching staff and the observer, in the total evaluation process, must be carefully defined. This individual should be present for the entire course or, if this is not possible, for selected days throughout the course. The independent observer should work closely with the teaching staff and he or she must have a clear understanding of the course objectives and how the lecturers are attempting to achieve these.

5. The senior instructor should have the responsibility for designing pre-course and post-course questionnaires. The former will be used to determine the motivations of students and what they hope to gain from the course. The latter will be the major instrument used in the summative evaluation of the course. If test questions are used to assess the learning achieved, these questions can be incorporated into the pre-course and post-course questionnaires.

6. The senior instructor may be given the responsibility for designing a follow-up questionnaire to examine behavioural changes in students. This questionnaire, which is probably best designed after the course is completed, should not be sent out until 12 months have elapsed from the conclusion of the course.

7. The senior instructor will prepare a comprehensive report on the results of the evaluation. This report will present the evaluation data, together with the instructor's interpretation and suggestions or recommendations to the organizers. The evaluation prepared by the independent observer may be included as part of this report. The instructor should certainly get an advance copy of the observer's account and should be allowed to comment on it in his own report.

8. The preparation of the final report will be the last task required of the instructor in the fulfilment of his contractual obligations. However, the instructor may be asked to participate further in the follow-up study on behavioural effects and to prepare a supplementary report on the results of this survey.

9. (In some cases, if the organizers feel it to be desirable, an independent evaluator can be employed from the outset. Under these conditions, the evaluator performs many of the tasks assigned to the instructor in the preceding list).

7. EVALUATING COST-EFFECTIVENESS

These guidelines have largely been restricted to the evaluation of an educational programme in terms of its effectiveness (as defined in 0.2). Cost-effectiveness studies in this field are less frequently performed. Such a study would normally be conducted in order to compare the effectiveness and costs of alternative approaches to achieving some desired educational objective. Given three approaches, all shown to be equally effective, the cheapest alternative will be the most cost-effective. Given three approaches all costing the same amount, the most effective approach (in terms of achieving the desired objectives) will also be the most cost-effective.

A rare example of a cost-effectiveness analysis is given by Grotelueschen et al[24]. Four alternative instructional approaches were compared in terms of their likely effectiveness and feasibility or ease of implementation (Exhibit 26). The costs, including unit costs per student, were then computed for all four approaches (Exhibit 27). If all four methods can produce roughly equivalent results (i.e., they are equally effective), the second approach is the most cost-effective, at least until the number of students per term reaches 400, when the fourth approach becomes more cost-effective. This example is put forward as a useful model for a cost-effectiveness analysis relating to educational programmes.

Wolf[75] and Hallack[25] have discussed the analysis of costs associated with educational programmes. Wolf also presents some general guidance on cost-effectiveness and cost-benefit analysis.

Exhibit 26 (a)

Effectiveness Considerations Relating to Four Instructional Alternatives

Reproduced from A. D. Grotelueschen et al. Evaluation in Adult Basic Education.[24] Copyright 1976, Interstate Printers and Publishers Inc. Reproduced by permission of the publisher.

	Traditional Small Group Instruction	Conventional Large Group Lecture-Lab	Individualized Instruction	
			Locally Developed Materials	Commercially Developed Materials
Effectiveness				
Posttest Achievement	Depends on instructional strategy, materials, and management system; could be high if a mastery learning model were used	Not as high as the individualized course which uses a mastery model	Significantly higher than the conventional course	
Unit Test Achievement	Could be high if a mastery learning model were used; this would be difficult to manage in this mode	Probably not as high as the individualized course which uses a mastery model; a mastery model here would be difficult to manage given the usual manpower resources	Average unit performance was 90% or higher for 3 successive quarters	
Student Satisfaction	Generally highly positive	Least positive of all the alternatives	Generally highly positive	
Efficiency	Average in-class time is 42 hours per term	Average in-class time is 42 hours per term	Average in-class time for self-paced students is 32 hours per term	

Exhibit 26 (b)

Feasibility Considerations Relating to Four Instructional Alternatives

	Traditional Small Group Instruction	Conventional Large Group Lecture-Lab	Individualized Instruction: Locally Developed Materials	Individualized Instruction: Commercially Developed Materials
Feasibility				
Instructor Receptivity	Highly receptive; most aspire to teach small, intimate classes	Receptivity varies; some like the lecturer role; student-oriented instructors dislike lecturing to large groups	Few are willing to devote the extra time to develop the required system and course materials; changes instructor's role to one of manager of learning	Few are willing to adopt or adapt complete systems developed by others; more would be willing to adopt than to develop an entire course; changes instructor's role to manager of learning
Student Receptivity	Highly receptive even though most 100-level teachers are lower ranked faculty	The expected mode for 100-level lecture-lab courses; reactions vary according to student's preferences	Students who enroll are receptive; enrollment data suggest that the course is perceived as too time consuming even though the result is a high grade	
Conditions for Learning	Tends to be instructor paced and controlled; students proceed with group regardless of performance; some opportunities for individualization in pacing and feedback	Large group requires it be instructor paced and controlled; students proceed with group regardless of performance; difficult to provide opportunities for individualization in packaging and feedback	Materials learner controlled; progress is individually paced; mastery learning approach provides multiple trials to demonstrate proficiency and receive optimum feedback	
Facility Requirements (assumes enrollment of 200)	Seven classrooms, 30 seats, one hour, three days per week; seven lab sections, 30 stations, two hours per section one day per week	A lecture room (200 seats), one hour, three days per week; seven lab sections, two hours per section one day per week	A lecture room (200 seats), one hour, one day per week, an individualized learning lab with 14 carrels to handle 200 students if open 50 hours per week	
	Present lab space can accommodate 25% of enrollees; incremental costs required for construction if lab work were required of everyone		Facilities not a problem because integration of lab and lecture permits flexible scheduling thus optimal facilities use	

Exhibit 27

Cost Analysis of the Four Alternatives Identified in Exhibit 26

Reproduced from A. D. Grotelueschen et al. Evaluation in Adult Basic Education.[24] Copyright 1976, Interstate Printers and Publishers, Inc. Reproduced by permission of the publisher.

	Traditional Small Group Instruction	Conventional Large Group Lecture-Lab	Individualized Instruction: Locally Developed Materials	Individualized Instruction: Commercially Developed Materials
Costs[a]				
R & D Investment, Replacement[b]			$ 49,059	$ 13,385
Operation: Per Term[c]	$ 8,701	$ 3,043	$ 3,804	$ 3,804
Operation: Ten Years[c]	$348,040	$121,720	$152,152	$152,152
Total Dollar Costs: Ten Years	$348,040	$121,720	$201,211	$165,537
Unit Costs[d]				
50 students per term	$ 15.25	$ 18.67	$ 25.15	$ 20.69
100 students per term	$ 15.30	$ 9.36	$ 12.57	$ 10.35
200 students per term	$ 13.38	$ 4.68	$ 6.29	$ 5.18
400 students per term	$ 13.48	$ 3.30	$ 3.14	$ 2.59

[a] All direct costs inflated 5% per year for ten years.

[b] These are the existing operational modes; R&D, investment, etc., are assumed to be nonincremental sunk costs.

[c] Operation costs are based on a student enrollment of 200 which is the capacity of present lecture hall.

[d] Unit costs include operation, R&D, investment, and replacement costs; depreciation is based on a ten year schedule.

8. REPORTING

The final task in an evaluation programme will usually be the preparation of a written report addressed to the body requiring the evaluation, probably the organization sponsoring or funding the course. Hopefully, the report will be more than a mere "signing off" on the project; it should include data and recommendations useful to the organizers in planning and implementing future activities. As Kinsey[31] points out, however:

> All too often the main result of education program evaluation is a report that finds a resting place on office shelves. (Page 1).

The better the report, presumably, the more likely it is to be used by the body to whom it is addressed; hence the importance of good reporting in this field.

Datta[10] and Leithwood et al.[38] have discussed the problems involved in successfully communicating evaluation results to decision makers. Datta presents some practical tips on this subject, adapted from an earlier report by Lee and Holley[37]. Because these are of general utility, they are reproduced in Exhibit 28. In some cases, an evaluation may be performed primarily to answer a single question (e.g., should this course be adapted for videotape presentation to secure a wider audience?). Under these circumstances, the reporter should get directly to the substance - re-stating the question, presenting (whenever possible, in tabular or graphical form) the data that bear on this question, and giving an interpretation and recommendations where appropriate. Exhibit 29, from Lee and Holley[37], gives an example of such a direct format.

Nadler[46] identifies three potential audiences for an evaluation report: (1) the sponsors (who may need results for future planning or for "marketing" a similar course in the future), (2) the students, who may welcome the opportunity to compare their own perceptions with those of the group as a whole, and (3) the instructors, who may want feedback in order to improve their own performance. Clearly, the report must be tailored to the type of audience for whom it is intended and must reflect the purpose for which the evaluation was performed.

Guidelines on reporting are included in several publications on evaluation (e.g., Grotelueschen et al.[24], Patton[48], Udinsky et al.[70] and Wolf[75]). The book by Patton is especially useful for its insightful discussion on the analysis and interpretation of evaluation results.

Exhibit 28

Practical Suggestions on Reporting Evaluation Results

From Datta[10] as adapted from Lee and Holley[37]
Reproduced by permission of The Johns Hopkins University Press.

Evaluation Audiences

1. Know your audiences
2. Find out what information they need and, if possible, when they need it
3. Try to understand each audience's viewpoint

Evaluation Message

4. Relate the information to action that must be taken
5. Do not give the audience more than it needs
6. Start with the most important information
7. Highlight the important points
8. Make your report readable
   a. Create an imaginary reader and keep this person in mind as you work
   b. Imagine that you are explaining the various aspects of the program to an interested but relatively uninformed person
   c. Check the vocabulary to make sure you have used familiar words
   d. Use action verbs as much as possible
   e. Cut out the deadwood
   f. Shorten your sentences
   g. Write shorter paragraphs
   h. Personalize your text
      (1) Use personal pronouns
      (2) Use contractions
      (3) Use "shirtsleeves" language

Verbal Presentations

9. Make the presentation format interesting and varied
10. Do what comes naturally
11. Make the visuals large and simple
12. Involve the audience in the presentation

Exhibit 29

Concise Format for Presenting Evaluation Data
Designed to Answer a Single Question

Reproduced from Lee, A. M. and Holley, F. M. Communicating evaluation information. (In: Morris, L. L. and Fitz-Gibbon, C. T. How to Present an Evaluation Report, pp. 27-47). Copyright 1978, Sage Publications Inc. Reproduced by permission of the publisher.

Decision Question

Should the district provide a fourth (summer) quarter for students in grades 7-12?

Information Summary

11% of the secondary students are interested in attending an additional fourth quarter, and 18% of the teachers are willing to teach an additional summer quarter.

All secondary students were surveyed about their reasons for wanting to attend a fourth summer quarter. The results:

Reasons given by students who said they would attend a summer quarter	Grades 7	8	9	10	11
For enrichment	26%	20%	19%	18%	27%
To graduate early	29%	22%	32%	45%	60%
To catch up	43%	57%	48%	34%	12%
Other	2%	1%	1%	3%	1%

All secondary teachers in the district were surveyed to assess their availability for teaching the fourth summer quarter. The results:

18%—would teach all four quarters
45%—could teach summer quarter if they could get off another quarter
33%—not available to teach a summer quarter under any circumstances
4%—want more information
100%

9. RELIABILITY AND RELATED CONCERNS

One important matter, alluded to earlier but not elaborated on, is that of the adequacy of the evaluation information that is collected. Knox[34] identifies five facets of "adequacy" in this context:

1. Validity
2. Reliability
3. Universality
4. Application of results
5. Efficiency

The adequacy of evaluation data is also discussed by Siegel et al.[56] who, in addition, provide a very useful review of various approaches to the design of achievement tests.

In terms of validity the important consideration is to ensure that the data collected do in fact tell us what we want to know. This implies that the evaluation, as well as the educational offering itself, must have defined objectives. That is, one must first carefully specify what it is one wants to learn about the course and then design evaluation procedures that will gather the data needed to draw appropriate conclusions. One must be certain that the evaluation procedures, collectively, gather all the data needed and, equally important, that they are not collecting superfluous data which will have no particular use. Also, one must be sure that the questions asked are questions that the respondents are competent to answer. When looking for completely frank responses, one may be more likely to get these if questionnaires are completed anonymously. Likewise, the anonymous approach may be preferable in any tests of learning acquired by students. A student may feel less threatened if not required to put his name on a test of this kind.

Reliability relates to the reproducibility of evaluation results. A questionnaire, or other evaluation instrument, may be considered reliable if it gives approximately the same results however many times it is applied in an identical situation. One could test the reliability of an evaluation questionnaire by applying it twice under exactly the same conditions (i.e., using the same group of students at exactly the same point in a course). This could be done, for example, by having the questionnaire completed before and after lunch on a particular day, but this does not seem a very practical or desirable approach. Usually we must be satisfied with less. At the very least, one must do everything one can to ensure that the questions are quite clear and that they are not susceptible to various interpretations. Closed-ended questions (i.e., those requiring a selection from a list of alternatives) are likely to get more consistent responses than open-ended questions. Standard scales, such as the Kropp-Verner scale mentioned earlier, are likely to produce quite reliable responses because these have been tested and proven with many previous audiences. It is desirable that a questionnaire be pretested before a course by asking various people, having roughly the same background as the actual participants, for their interpretation of the meaning of the questions. A question that is always interpreted in the same way, by members of this pretest group, may be regarded as unambiguous and reliable. A question that is variously interpreted will require modification for purposes of clarification. As more courses are held by a particular organization, and more evaluations are conducted, it will be possible to identify questions or types of questions that appear to produce reliable responses. Wherever possible, it seems desirable to develop fairly standardized evaluation instruments - questionnaires that can be

applied, perhaps with minor modification, to several courses. This has two advantages: (a) reliability is improved by the use of questions that have been tested in the past, and (b) the impact of various courses can be compared if the evaluation methodology is held constant.

This brings us logically to the next characteristic, that of universality. Each educational offering within a particular programme will have some unique features that one may want to evaluate. But the various programmes will also have many things in common. If the same approach is used to the evaluation of the common elements it will be possible, as Knox[34] has pointed out, to combine findings in order to develop improved guidelines for the planning and implementation of courses. Analysis of responses to common elements would allow the planning body to identify characteristics of courses or participants that, in the words of Knox, "are associated with high satisfaction or high achievement and those associated with low satisfaction or achievement". It is strongly recommended that a sponsoring organization should encourage the development of standard approaches to the evaluation of courses, workshops and seminars. Each set of evaluation instruments should include both common elements and elements that may be needed to evaluate the unique features of a particular programme.

The characteristic application of results obviously relates to the use that is to be made of the evaluation data. One must be quite sure to collect data in such a way that they can be applied to answer all the questions that one would like to answer about a particular programme. To take a simple example, one may want to know what effect the amount of previous professional experience has on the participants' reaction to a course or on the degree to which they seem to have benefited from it. Quite obviously, if forms are completed anonymously, one must ask respondents to indicate the number of years they have been active in the field. In this way one can separate the responses of the more experienced participants from those of less experience. This highlights once more the need to prepare evaluation specifications; that is, to define very carefully in advance what it is that one wants to learn about a particular course. It seems extremely important, then, that those responsible for planning a course should prepare a set of specific questions for which they would like to obtain answers. The evaluation should then be designed to answer these questions as reliably, thoroughly and efficiently as possible.

Efficiency relates to the way data are collected. It is the responsibility of the evaluator to see that the needed data be collected in the easiest possible way from those in the best position to provide it. The evaluator must avoid redundancy: he must not collect data that he will not use and he must not collect the same data twice (unless he feels he needs to as a check on reliability). Questionnaires should not be used to gather data that are readily available in other ways (e.g., the application forms of participants). The data-gathering procedures should interfere as little as possible with the conduct of educational programmes. One must avoid evaluation "overkill" at all costs. The evaluator must consider the best time to collect various data as well as the best way of collecting it. It is also important to plan the data collection in such a way that reduction, analysis, and interpretation of the data are facilitated. Because many short courses will involve relatively small numbers of participants, it is likely that the data reduction can be handled quite adequately by hand. If very large quantities of data are to be collected, however, serious consideration should be given to the possibility of collecting these data in such a way that they may be manipulated by standard data processing techniques. The report by Whitney and Brandenburg[73] contains useful guidelines on the design of questionnaires to facilitate subsequent data processing operations.

10. CONCLUSION

This report has attempted to provide guidelines that can be used in the evaluation of courses, workshops, seminars, and similar educational endeavours. It deliberately restricts itself to the presentation of guidelines rather than trying to lay down rigid rules or "standards" for the conduct of evaluation programmes. The guidelines relate to what data to collect and how, in general, they may be collected. They are illustrated with examples of questionnaires, tests and reports. This is not, however, a cookbook approach; recipes to be slavishly followed are not here. On the other hand, the report does present various ingredients from which a recipe for a particular evaluation could be put together. Although, as stated several times already, it is desirable to standardize the collection of data representing elements common to several courses, there is also considerable room for experimentation and innovation in the evaluation of individual offerings within a particular programme. Such experimentation and innovation should be encouraged.

It should be obvious, of course, that not all of the procedures described in these guidelines would be applied in any one situation. That would constitute "evaluation overkill" on a grand scale.

The dangers of this have been emphasized by Woodward and Yeager[76]. If evaluation procedures are too extensive and/or redundant, the evaluation process may interfere with the instruction itself. Moreover, if too much evaluation is attempted, the results may be unsatisfactory. Wolf[75] cautions:

> Don't dilute the evaluation effort by trying to measure everything and end up doing a mediocre or even poor job. (Page 30).

Woodward and Yeager also stress the need for flexibility in the approach to evaluation. One must be prepared to modify evaluation procedures if they seem not to be giving the information needed or if one can identify alternative, possibly better procedures as the course progresses. This is particularly relevant, of course, to formative evaluation.

> Formative evaluation must be 'adaptive' evaluation. This implies that the evaluation design and procedures must be sufficiently flexible to adapt to the changing requirements of the particular evaluation context in which the evaluation is being conducted, and the evaluation model must allow for such flexibility. (Page 13).

There are cost-effectiveness considerations relating to evaluation itself. How much evaluation is performed, and how sophisticated the evaluation procedures adopted, must be related to the size, complexity, and costs of the educational programme itself. An extensive and expensive programme of courses, applied worldwide to agricultural extension agents or family planning counsellors, may warrant an elaborate and expensive evaluation, perhaps exploring all the facets identified in 0.2. On the other hand, a single course with few participants will usually merit only a simple and inexpensive approach.

Farmer[17] has put this aspect in perspective:

> A justifiable evaluation effort must also be cost effective. The money and time spent on a thorough evaluation of all aspects of a program could exceed that required to conduct the program itself. (Page 29.)

Above all, evaluation in education, as in other fields, should be performed with practical objectives in mind. The evaluation should not become an end in itself. The proliferation of evaluation models and approaches, as described in the literature, suggests that evaluation in education has become an end in itself, at least in some quarters. Knowles[33] seems to recognize this when he refers to evaluation becoming a "sacred cow" and much overemphasized.

An evaluation need not be a "model" of experimental design. Davis and McCallon[11] suggest, in fact, that too much "scientism" may have crept into the field, brought in by academicians who:

> ... have a strong need for scientific measurements, for pre- and post-testing, control groups and the like. Aside from the near impossibility of achieving adequate controls and the expense and time involved, I have no objection to this kind of precision, but it simply isn't needed in the daily practice of adult education. (Page 277.)

A similar point has been made by Giuliano and Jones[19] in a completely different context:

> One decision which has to be made is whether the effort is to be put into proof-oriented experiments or into insight-oriented experiments. By the former, we mean concentration of evaluator effort into large-scale systematic and statistically valid investigations of one type of variability with other conditions being relatively fixed; by the latter we mean a spreading of evaluator effort over a number of investigative forays designed to give not proof but insight as to how the experimental variables interact with one another.
>
> A proof-oriented experiment should lead to a well-defined statement of conclusion backed up with an analysis of variance of the results and identified confidence limits. However, such experiments are premature unless one knows exactly what one wants to prove and the conditions under which the proof is interesting. It is not of much interest to know that search option A is proven to be better than search option B under given conditions with confidence 0.999 - what is really of interest is whether the given conditions are actually realistic, how much better is A than B, and is this better enough to be of real concern.
>
> Insight-oriented experiments may or may not lead to well-defined conclusions, and one such experiment may or may not be sufficiently meaningful statistically to constitute convincing proof in the face of withering doubt. However, several such tests can sometimes be performed for the cost of one proof-oriented test, and the pattern of observed results might tell a lot more about the system being investigated than any single test, no matter how firm the conclusions of that one test are. (Page 43).

This philosophy is also supported by Fischer[18] in an interesting discussion of the economics of statistical sampling:

> Statistics can be an extremely valuable tool to extract the most information from given data, but it cannot replace expert judgment where that is required. An expert's intuition is still irreplaceable whenever it is necessary to combine and interpret a very large number of hints and pieces of information which are not easily quantifiable, and where delicate judgments must be made. Such situations are often found in the social sciences, but also for example in medicine. There the most sophisticated tools don't help much if they are not used by an experienced person. Someone once said that if he had to make a choice, he would rather be operated on by a top surgeon using a rusty kitchen knife, than by a manufacturer of surgical instruments using his latest equipment. This is a somewhat macabre formulation, but to the point. Of course it does not mean that a good surgeon using modern instruments is not preferable still. What is needed is communication and cooperation between experts in different fields. (Page 338).

Finally, Patton[48] has argued the merits of "qualitative evaluation". While the quantitative experimental methods of the natural sciences have their place, even in social science applications, the "scientific method" is not the only way in which reliable evaluation data can be gathered. Patton describes how qualitative methods can also produce useful and valid data.

Although an evaluation need not be a masterpiece of research, it should produce results that one can place some reliance on. That is, the results need not prove anything, in a statistical sense, but they should at least provide some clear insight on what has occurred and why. One must therefore try to gather the best data possible (qualitative, quantitative, or both) within the resources available. It is hoped that these guidelines may provide assistance in this.

Since education and training programmes may have very long-term effects on participants, the impact of such programmes cannot be fully assessed by any evaluation procedures. As Cooley and Lohnes[9] have pointed out:

> Evaluation is defined as a process rather than a product. Educational procedures are never completely, finally evaluated. (Page 3).

REFERENCES CITED

(The reports identified by ED numbers in this bibliography are available, in paper copy or microfiche, from: EDRS, P.O. Box 190, Arlington, Va. 22210, U.S.A)

1. Ahmann, S.; Glock, M.D. Evaluating Pupil Growth: Principles of Tests and Measurements. 4th ed. Boston, Allyn and Bacon, 1971, 621 p.

2. Bloom, B.S.; Hastings, J.T.: Madaus, G.F. Handbook on Formative and Summative Evaluation of Student Learning. New York, McGraw-Hill, 1971, x, 923 p.

3. Braskamp, L.A. et al. Guidebook for Evaluating Teaching. Urbana, University of Illinois, Office of Instructional Resources, 1983, 71 p.

4. Bryant, V.E. An evaluation of continuing education programs based on the principles of adult learning. In: The Evaluation of Continuing Education for Professionals: a Systems View, pp. 326-334. Seattle, University of Washington, 1979, ix, 416 p.

5. Burstiner, I. Evaluation of a mini-workshop in "creative leadership" for educational administrators. Education 93, 1972: 47-50.

6. Campbell, M.D. Before You Plan Educational Programs: Find Out What's Needed. Urbana, University of Illinois, Office of Continuing Education and Public Service, 1980, 22 p.

7. Carmichael, J.; Tanner, L.; Carmichael, L. Research in human sexuality education. Medical Education 11 (2), 1977: 111-113.

8. Chang, H.C. Narrative Evaluation Report on the Institute for Training in Librarianship. St. Croix, College of the Virgin Islands, 1976, 133 p.

9. Cooley, W.W.; Lohnes, P.R. Evaluation Research in Education. New York, Irvington Publishers, 1976, xii, 368 p.

10. Datta, L.-E. Communicating evaluation results for policy decision making. In: Berk, R.A., ed. Educational Evaluation Methodology: the State of the Art, pp. 124-145. Baltimore, The Johns Hopkins University Press, 1981, xii, 168 p.

11. Davis, L.N.; McCallon, E. Planning, Conducting and Evaluating Workshops. Austin, Texas, Living Concepts, 1974, xii, 310 p.

12. Dixon, J.K. Methodological Considerations in Evaluation of Continuing Education in the Health Professions. Paper presented at the Annual Convention of the American Educational Research Association, New York, 1977, 27 p. ED 138780.

13. Dizney, H. Classroom Evaluation for Teachers. Dubuque, Iowa, W.C. Brown, Co., 1971, xii, 119 p.

14. Edwards, A. Techniques of Attitude Scale Construction. New York, Appleton-Century-Crofts, 1957, 256 p.

15. Evaluation Guidelines. Carson City, Nevada, State Department of Education, July 1973, 75 p. ED 086727.

16. Evaluation of Adult Education Programs. Sacramento, California State Department of Education, 1979, 12 p. ED 171980.

17. Farmer, J.A., Jr. Evaluation of broad-aimed education programs. In: Program Evaluation, pp. 23-45. New York, World Education, 1975, 68 p. ED 120394.

18. Fischer, D. On the economics of statistical sampling. In: Machlup, F., et al. Information Through the Printed Word. Vol. 2, pp. 319-338. New York, Praeger, 1978, xxiii, 338 p.

19. Giuliano, V.E.; Jones, P.E. Study and Test of a Methodology for Laboratory Evaluation of Message Retrieval Systems. Cambridge, Mass., Arthur D. Little Inc. 1966, 183 p.

20. Good, C.V. Dictionary of Education. 3rd ed. New York, McGraw Hill, 1973, 495 p.

21. Greenfest, M.; Bragg, S. A Methodology and Design for Evaluating Short-Term Training, 1976, 45 p. ED 122023

22. Gronlund, N.E. Constructing Achievement Tests. Englewood Cliffs, New Jersey, Prentice-Hall, 1968. iii, 118 p.

23. Gronlund, N.E. Measurement and Evaluation in Teaching. New York, Macmillan, 1965, xiii, 420 p.

24. Grotelueschen, A.D.; Gooler, D.D.; Knox, A.B. Evaluation in Adult Basic Education: How and Why. Danville, Ill., Interstate Printers and Publishers Inc., 1976, xii, 274 p.

25. Hallak, J. Cost analysis in evaluating educational programs. In: Lewy, A.; Nevo, D. (eds.) Evaluation Roles in Education, pp. 475-486. London, Gordon and Breach, 1981, xii, 512 p.

26. Hampton, L.A. Evaluating continuing education programs. Adult Leadership 22 (3) September 1973: 105-119.

27. Houge, D.R. Evaluation by Participant/Observers. 1981, 10 p. ED 206252.

28. Isaac, S.; Michael, W.B. Handbook in Research and Evaluation. San Diego, R.R. Knapp, 1971, vi, 186 p.

29. Jaster, F. Assessing Corporate Training Programs in Business Communications, 1981, 16 p. ED 209709

30. Joint Committee on Standards for Educational Evaluation. Standards for Evaluations of Educational Programs, Projects, and Materials. New York, McGraw Hill, 1981, xxvi, 161 p.

31. Kinsey, D.C. Evaluation in Nonformal Education: the Need for Practitioner Evaluation. Amherst, University of Massachusetts, Center for International Education, 1978, v.29 p.

32. Kirkpatrick, D.L. Evaluation of training, In: Craig, R.L.; Bittel, L.R. (eds.). Training and Development Handbook, pp 87-112. New York, McGraw-Hill, 1967, xii, 650 p.

33. Knowles, M.S. The Modern Practice of Adult Education. New York, Association Press, 1970, 384 p.

34. Knox, A.B. Approaches to Conference Evaluation. Working paper prepared for the Conference and Institute Division Meeting, National University Extension Association, Lincoln, Nebraska, 29 April-2 May 1962. Available from the Office for the Study of Continuing Professional Education, University of Illinois, Urbana, Ill.

35. Knox, A.B. Program Evaluation in Adult Basic Education. Tallahassee, Florida State University, 1971, 21 p.

36. Kropp, R.P.; Verner, C. An attitude scale technique for evaluating meetings, Adult Education 7, 1957: 212-215.

37. Lee, A.M.; Holley, F.M. Communicating evaluation information: some practical tips. In: Morris, L.L.; Fitz-Gibbon, C.T. How to Present an Evaluation Report, pp. 27-47. Beverly Hills, Calif., Sage Publications Inc., 1978, 80 p.

38. Leithwood, K.A.; Wilson, R.J.; Marshall, A.R. Increasing the influence of evaluation studies on program decision-making. In: Lewy, A.; Nevo, D. (eds.) Evaluation Roles in Education, pp. 357-383. London, Gordon, and Breach, 1981, xii, 512 p.

39. Lindquist, E.F. (ed.). Educational Measurement. Washington, D.C., American Council on Education, 1951, xix, 819 p.

40. Linn, R.L. Measuring pretest-posttest performance changes. In: Berk, R.A., ed. Educational Evaluation Methodology: the State of the Art, pp. 84-109. Baltimore, The Johns Hopkins University Press, 1981, xii, 168 p.

41. McGuire, C.H.; Babbott, D. Simulation techniques in the measurement of problem solving skills. Journal of Educational Measurement 4 (1) Spring 1967: 1-10.

42. McManis Associates Inc. Evaluation of LEAA Funded Courts Training Programs. Volume II. Washington, D.C., U.S. Department of Justice, Law Enforcement Assistance Administration, 1980, 357 p.

43. Mehlinger, H.D.; Patrick, J.J. The Use of "Formative" and "Summative" Evaluation in an Experimental Curriculum Project: a Case in the Practice of Instructional Materials Evaluation. Bloomington, Indiana University, High School Curriculum Center in Government, March 1970, 16 p. ED 041443.

44. Meyer, G.R.; Jenkins, C. Preliminary studies of the effectiveness of minicourses for the in-service education of teachers and trainers. Programmed Learning & Educational Technology 16 (3) August 1979: 210-218.

45. Misanchuk, E.R. Uses and Abuses in Continuing Education Programs. Paper presented at the Adult Education Research Conference, San Antonio, Texas, 1978, 22 p. ED 160734.

46. Nadler, L. Improving the results of workshops. Part 3. Linkage, evaluation and follow-up. Training and Development Journal 30, (9) September 1976: 31-35.

47. Oppenheim, A.N. Questionnaire Design and Attitude Measurement. New York, Basic Books, 1966, ix, 208 p.

48. Patton, M.Q. Qualitative Evaluation Methods. Beverly Hills, Calif., Sage Publications, 1980, 379 p.

49. Payne, S.L. The Art of Asking Questions. Princeton, Princeton University Press, 1951, xiv, 349 p.

50. Penn, P.D. Project Born Free: Evaluation of the 1978 Born Free National Institute. Minneapolis, University of Minnesota, College of Education, 1978, 197 p. ED 193595.

51. Postlethwait, S.N. et al. The Evaluation of Minicourses in Undergraduate Biology. Lafayette, Indiana, Purdue University, 1974, 20 p. ED 193008.

52. Provus, M. Evaluation of ongoing programs in the public school system. In: Tyler, R.W. (ed.). Educational Evaluation: New Roles, New Means, pp. 242-283. Chicago, University of Chicago Press, 1969. xiii, 409 p.

53. Roid, G.H. Systems Design for Course Evaluation. Paper presented at the AERA Symposium "Research, Systems Design and the Future of University Course Evaluation", Chicago, April 1972, 43 p. ED 064398.

54. Scriven, M.S. The methodology of evaluation. In: Tyler, R.W. et al (eds.). Perspectives of Curriculum Evaluation, pp. 39-83. Chicago, Rand McNally, 1967, 102 p. (AERA Monograph Series on Curriculum Evaluation, 1).

55. Shaw, M.E.; Wright, J.M. Scales for the Measurement of Attitudes. New York, McGraw-Hill, 1964, xxii, 604 p.

56. Siegel, A.I.; Bergman, B.A.; Federman, P. Some Techniques for the Evaluation of Technical Training Courses and Students. Wayne, Pa., Applied Psychological Services, Inc., February 1972, 137 p. AD 753094.

57. Silver, P.F. The Effects of Attitudes on Quality Judgments of Inservice Programs. Paper presented at the Annual Meeting of the American Educational Research Assocation, 1981, 15 p. ED 203533.

58. Slater, M. Short Course Assessment and Evaluation in the Library/Information Field. London, Aslib, 1974, 66 p. (Aslib Occasional Publication No. 15).

59. Smith, G.D. Directing a Workshop for Implementers of Career Education. Greeneville, Tennessee, Greeneville City Schools, 1974, 26 p. ED 107129

60. Stake, R.E. The countenance of educational evaluation. Teachers College Record 68 (7) April 1967: 523-540.

61. Stake, R.E. Objectives, priorities and other judgment data. Review of Educational Research 40 (2) April 1970: 181-212.

62. Stake, R.E. A research rationale for EPIE. The EPIE Forum 1 (1) September 1967: 7-15.

63. Stake, R.E. Responsive Evaluation. December 1972, 4 p. ED 075487.

64. Stake, R.E. et al. Evaluating a Regional Environmental Learning System: a Program Evaluation Manual. Urbana, University of Illinois, College of Education, 1979, 148 p.

65. Steele, S.M. Contemporary Approaches to Program Evaluation and Their Implications for Evaluating Programs for Disadvantaged Adults. Syracuse, N.Y., ERIC Clearinghouse on Adult Education, 1973, 257 p.

66. Studebaker, D.P. et al. Evaluation Training Materials: the State of the Art. 1979, 56 p. ED 177224.

67. Stufflebeam, D.L. Metaevaluation: concepts, standards and uses. In: Berk, R.A., ed. Educational Evaluation Methodology: the State of the Art, pp. 146-163. Baltimore, The Johns Hopkins University Press, 1981, xii, 168 p.

68. Stufflebeam, D.L. The use of experimental design in educational evaluation. Journal of Educational Measurement 8, 1971: 267-274.

69. Thiede, W. Evaluation and adult education. In: Jensen, G. et al (eds.). Adult Education: Outlines of an Emerging Field of University Study, pp. 291-305. Washington, D.C., Adult Education Association of the U.S.A., 1964, xiv, 334 p.

70. Udinsky, B.F.; Osterlind, S.J.; Lynch, S.W. Evaluation Resource Handbook. San Diego, Calif., EdITS Publishers, 1981. iv, 250 p.

71. U.S. Civil Service Commission. An Evaluation Model Designed to Measure the Effectiveness of a One-week Training Course "Position Classification and the Management Process". Washington, D.C., June 1970, 48 p. ED 064468.

72. Veri, C.C. et al Evaluating Teaching Effectiveness in Extension-Administered Programs. St. Louis, University of Missouri, Extension Division, 1973, 15 p. ED 078275.

73. Whitney, D.R.; Brandenburg. D.C. The Questionnaire as a Data Source. Urbana, University of Illinois, Office of Instructional Resources, August 1974, 45 p.

74. Wilsing, W.C. Program development: transforming needs into objectives. In: The Evaluation of Continuing Education for Professionals: a Systems View, pp. 242-264. Seattle, University of Washington, 1979, ix, 416 p.

75. Wolf, R.M. Evaluation in Education. New York, Praeger, 1979, xiv, 217 p.

76. Woodward, J.P.; Yeager, J.L. Evaluation of Programs to Train Educational R & D Personnel. Pittsburgh, University of Pittsburgh, Learning Research and Development Center, April 1972, 26 p. ED 064376.

APPENDIX 1

Examples of Questionnaires Used in the
Ongoing Evaluation of a Course or Workshop

(A) Daily reaction form used in University of Illinois workshop on "Education and the Elderly". Reproduced by permission of the University of Illinois.

_____
Identification Number

DAILY REACTION FORM
Education and the Elderly

(The items on this form refer to workshop activities during the past 24 hours.)

1. The following three items refer to the workshop sessions in which all participants met as a total group.

   a. Check the one phrase that best indicates how <u>important</u> and relevant the subject-matter and discussion topic was for you.

      5 ( ) Most important
      4 ( ) Very important
      3 ( ) Some importance
      2 ( ) Little importance
      1 ( ) No importance

   b. Check the one phrase that best indicates the extent to which the presentations had <u>clear</u> and to-the-point explanations.

      5 ( ) Exceptional
      4 ( ) Very good
      3 ( ) Good
      2 ( ) Fair
      1 ( ) Poor

   c. Check the one phrase that best indicates how understandable and useful the educational <u>materials</u> related to the total group sessions were.

      5 ( ) Exceptional
      4 ( ) Very good
      3 ( ) Good
      2 ( ) Fair
      1 ( ) Poor

2. Write in the following space the number of the small <u>sub-group</u> in which you are a participant.

   Group Number _____

3. Check the one phrase that best indicates how valuable your participation in this small sub-group has been for you.

   5 ( ) Exceptional
   4 ( ) Very good
   3 ( ) Good
   2 ( ) Fair
   1 ( ) Poor

4. Check the one phrase that best indicates how satisfactorily you are progressing on your project, reading, and independent study.

   5 ( ) Very well
   4 ( ) Reasonably well
   3 ( ) Somewhat
   2 ( ) Hardly
   1 ( ) Not at all

5. List below the most valuable new ideas related to the workshop that you have dealt with during the past 24 hours.

   _____
   _____
   _____
   _____

6. List below the main suggestions that you have which would have most improved your workshop experience during the past 24 hours.

   _____
   _____
   _____
   _____

7. Check the one phrase that best indicates your general rating of your workshop experience during the past 24 hours.

   5 ( ) Really outstanding
   4 ( ) Very satisfactory
   3 ( ) Average
   2 ( ) Just acceptable
   1 ( ) Needs improvement

8. List below any other comments you have about the workshop at this point that might help to improve this and future workshops.

   _____
   _____
   _____
   _____
   _____

1(B) Weekly reaction form used in UNISIST Course on Information Retrieval and Information Retrieval Systems, Katowice, Poland, August 1974.

Weekly Reaction Form

1. Check the one statement that best indicates how relevant the subject-matter of this week has been to you.

   Of great relevance ☐

   Of some relevance ☐

   Of little relevance ☐

   Of no relevance ☐

2. Check the one statement that best indicates your reaction to the quality of the lectures.

   Very good ☐

   Good ☐

   Average ☐

   Poor ☐

   Very Poor ☐

3. For each of the instructional methods mentioned below check the box that best describes your reaction to its value as a learning experience.

	Great value	Some value	Little value	No value
Transparencies				
MEDLARS slides				
Films				
MEDLINE demonstration				
Practical exercises				
Discussion periods				

4. Check the one statement that best describes how you feel you are progressing towards satisfying your own objectives in attending this course.

- Great progress ☐
- Good progress ☐
- A little progress ☐
- Hardly any progress ☐
- No progress at all ☐

5. During this week what specific aspects of the course were <u>most</u> valuable in achieving your own objectives in attending?

6. During this week what specific aspects of the course were <u>least</u> valuable in achieving your own objectives in attending?

7. List any suggestions you have that you think might improve the course in later days or might improve the course if given again in the future.

8. Check the one statement that best indicates your <u>general evaluation</u> of the course during this week.

- Very good ☐
- Good ☐
- Average ☐
- Poor ☐
- Very poor ☐

9. If you have been able to complete some of the recommended supplementary readings for the course please indicate how valuable these readings have been to you so far.

Very valuable ☐

Valuable ☐

Of little value ☐

Of no value ☐

10. Select the statement that best indicates the novelty of the material presented this past week.

Most of the material was new to me ☐

About half the material was new to me ☐

Less than half the material was new to me ☐

None of the material was new to me ☐

11. Do you think the material presented in the past week could have been better presented some other way?

☐ No

☐ Yes What suggestions do you have for improved presentation?

APPENDIX 2

Example of a Checklist for Use by an Independent Observer

Reproduced from: U.S. Civil Service Commission. Bureau of Training. Training Evaluation: a Guide to its Planning, Development, and Use in Agency Training Courses. Washington, D.C., May 1971. 21 p. Pamphlet T-13. ED 054 365.

CLASSROOM OBSERVATION GUIDE

Course _____ Session _____

Time _____ Date _____ Instructor _____

1. Training objectives: _____

2. Assessment: To what extent did training objectives appear to be attained? (Adverse factors should be explained in item 4 below.)

   a. ☐ Fully. (Minor deficiencies, if any.); b. ☐ Partially. (Some factors preventing full achievement.); c. ☐ Hardly at all.

3. Checks on attainment: What steps did the instructor take to assure himself that training objectives were being attained? _____

   How effective were they? _____

4. Factors influencing objective attainment: Factors that had an adverse effect on training objective attainment as indicated in 2 above. (Check any such factors and explain on the back of this sheet.)

Factor	Check
A. Introduction to training session	_____
B. Content	_____
C. Clarity of explanations	_____
D. Instructor's apparent knowledge of subject	_____
E. Subject-matter organization	_____
F. Selection and use of training aids	_____
G. Instructor's manner (tact, etc.)	_____
H. Instructor's ability to maintain trainees' interest	_____
I. Instructor's ability to direct discussion	_____
J. Appropriateness of material to group	_____
K. Time/material relationship	_____
L. Individual trainee involvement	_____
M. Instructional methods used	_____
N. Reference materials	_____
O. Classroom facilities	_____
P. Tests and practical exercises	_____
Q. Other (specify)	_____

APPENDIX 3

Interview Guide Used in Follow-up Telephone
Interviews with Participants in a Seminar on Writing Skills.

Reproduced from Jaster[29] by permission of the author.

DATE_____

NAME OF
INTERVIEWEE_____

Phone #_____

Introduction:

Hello, my name is _____ and I'm calling from _____ University. We're conducting a research survey of those who attended the company's Written Communications Seminar in 1978-79. The purpose of this survey is to get information which will help improve future workshops.

The interview will take about fifteen minutes. Is this a convenient time for you or should I call back?

(If call back: When would it be convenient for me to call you back?)

DATE_____

TIME_____

NOTE: Be sure to call back at the proper time.

1. I'd like to begin by asking you to name things about the workshop that you particularly liked or disliked:

	Volunteered		Probed	
	Like	Dislike	Like	Dislike
Teacher(s)				
Format (3 day)				
Setting (Location)				
Scheduling (M-W, W-F)				
Organization				
Content				
Other (Specify)				

2. What are the two or three most important things you learned from the workshop?

(IF MORE THAN ONE, ASK:)

Which of these is the most important?

Why do you think it is the most important thing you learned from the workshop?

3. What kinds of writing are part of your day-to-day activity?

A. Requests for funds _____

B. Job Orders _____

C. Requisitions _____

D. Instructions _____

E. Manuals _____

F. Reports: Type? _____

_____

G. Other: Specify _____

_____

_____

_____

4. This next question deals with the amount of change in your writing abilities that was brought about as a result of the workshop.

I'd like for you to picture in your mind a 5 point scale.

"1" represents little or no change; "5" represents a great deal of change.

With this scale in mind, please select a number on the scale which best represents the extent to which your writing skills have changed as a result of the workshop. Remember, "1" represents little or no change; "5" represents a lot of change.

5_____4_____3_____2_____1_____

NOTE: Ask: Why did you choose 1 or 5 only if E chooses 1 or 5.

5. In the next few questions, I'm going to ask you to rate the importance of writing skills in your job.

   I will ask you to tell me for each item whether you think it is Very Important, Important, or Not Very Important.

   How important is your writing skill to the effective performance of your job?

   Would you say it is:

   Very_____Important_____Not Very_____

   In your opinion, how important is being able to write effectively to success in your field?

   Very_____Important_____Not Very_____

   How important is it for good supervisors to write effectively?

   Very_____Important_____Not Very_____

6. In the workshop, the instructors covered several topics. Using the same scale, VI, I, NVI, please rate the following topics in terms of their importance to your writing:

How to organize your material	VI	I	NVI
How to forecast your content	VI	I	NVI
How to be concise	VI	I	NVI
How to be clear	VI	I	NVI
How to avoid passive constructions	VI	I	NVI
How to analyze the reader's needs	VI	I	NVI
Use of sample reports and formats as examples to follow	VI	I	NVI
Business Writer's Handbook as a desk reference	VI	I	NVI

7. How much of your daily time is spent in writing?

   More than 50%__; more than 25%__; more than 10%__; less than 10%__.

   What percent of your daily time at work is spent in other kinds of communications?

   Reading?_____; speaking/talking?_____; listening?_____

8. Although the workshop focused on writing skills, has what you learned in that class affected your abilities or skills in other areas?

   (If asked to explain: has what you learned there improved your ability, for example to read critically, that is, your ability to judge the quality of what others have written?)

   In what ways has your reading ability improved?

9. In the months since you took the workshop, have you referred to any of the materials you received from the instructors?

   Yes _____ No _____

   Which materials?

   About how many times a week/month have you referred to them?

   Have you referred to the Business Writer's Handbook?

   Yes _____ No _____

   About how often? weekly? _____ monthly? _____

10. If the company were to offer a one-day follow-up session, what topics would be more helpful to you?

11. What else about the workshop would you like to comment on?

12. Demographics

    Position/Title_____

    Have you been promoted within the last two years? Yes____ No ____

    Age: 20-30 _____
    30-40 _____
    40-50 _____
    50-60 _____
    Over 60 ___

    College degree? Yes____ No ____ Type of degree?_____

    Graduate or Professional degree? Yes____No____Type of degree_____

    DO NOT ASK, JUST RECORD:

    Sex: M_____ F_____

APPENDIX 4

Follow-up Questionnaire as Used in the Evaluation of Training Programmes Sponsored by the Law Enforcement Assistance Administration, U.S. Department of Justice.

Reproduced by permission from McManis Associates Inc./42

E. ASSESSMENT OF MOST RECENT TRAINING COURSE

Please assess training process provided by _____ in terms of its
(please fill in)
strengths and weaknesses. In your view, based on the most recent course taken at that institute:

(Please circle the appropriate number)

	Not at all	Some what	Adequate	Very Much	Fully
13. To what extent was the course applicable to the needs of the organization/system that sent you?	1	2	3	4	5

b. Further observations:

14. To what extent:

(Please circle the appropriate number)

	Not at All	With Reservations	Neutral	Fairly Willing	Extremely Interested
a. Did you want to attend training?	1	2	3	4	5
b. Was the training encouraged by the organization/system for which you work?	Hostile to the idea 1	Discouraging 2	Neutral 3	Very encouraging 4	It was mandatory 5

c. Further observations:

15 With regard to the general design of the course, to what degree:

	(Please circle the appropriate number)				
	Not at All	Some-what	Fairly Well	Very Much	Thoroughly
a. Was training based on the general needs of your profession?	1	2	3	4	5
b. Were your individual training needs taken into consideration during the course?	1	2	3	4	5
c. Were expected outcomes presented to you at the start of the course?	1	2	3	4	5
d. Were the learning objectives clear and succinct?	1	2	3	4	5
e. Was the achievement of the learning objectives something that could actually be demonstrated?	1	2	3	4	5
f. Did the training itself provide the opportunities to practice what you were taught?	1	2	3	4	5
g. Were you kept informed of your progress?	1	2	3	4	5
h. Did the training objectives, activities, and materials appear to fit together?	1	2	3	4	5
i. Did the instructor or other participants provide you with feedback on your performance that was useful to you?	1	2	3	4	5
j. Were you given the opportunity to provide input to the faculty during the course?	1	2	3	4	5

k. Further observations:

16. While individual instructors may vary, what did you think of the faculty in general with respect to:

	(Please circle the appropriate number)				
	Totally Unknow-ledgeable	Minimally Knowledge-able	Adequate	Quite Know-ledgeable	Expert
a. The degree of their substantive knowledge of the law/courts?	1	2	3	4	5
b. The extensiveness of their practical experience in the topic area?	1	2	3	4	5
c. Their teaching ability?	1	2	3	4	5

d. Further observations:

	(Please circle the appropriate number)				
	Not at all	Once or twice	Periodically	Often	Continually
17. How often have you called upon the staff/instructors since training for followup technical assistance to yourself and/or your jurisdiction?	1	2	3	4	5

Further observations:

18. Do you:

	(check one)	
a. Wish to return to the institute for more sessions in the future?	1 ☐ Yes 2 ☐ No	Why?
b. Attempt to make personal changes related to what you learned at the institute?	1 ☐ Yes 2 ☐ No	If so, please give examples
c. Attempt to make changes in your organization that were spurred by what you learned at the institute?	1 ☐ Yes 2 ☐ No	If so, please give examples
d. Share institute materials to assist others?	1 ☐ Yes 2 ☐ No	If so, which materials and for whom?
e. Recommend the institute to others?	1 ☐ Yes 2 ☐ No	If so, how and to whom? If not, why?

19. With regard to the training setting:

	(check one) Yes	No
a. Was the group of participants made up of people with similar professional roles and experience?	1 ☐	2 ☐
b. Was the number of participants in a class small enough to allow individual attention from the instructor?	1 ☐	2 ☐
c. Was the course long enough to allow you to meet the learning objectives?	1 ☐	2 ☐
d. Was there an effective match between your role/needs and the instructor's areas of expertise?	1 ☐	2 ☐
e. Were instructors available to you outside class?	1 ☐	2 ☐
f. Was the total number of contact hours you had with instructors adequate?	1 ☐	2 ☐
g. Were the training support services (e.g., copying, handouts, chalk board, graphics, etc.) adequate?	1 ☐	2 ☐
h. Were the physical accommodations (e.g., classrooms, refreshments, seminar rooms, lodgings) adequate?	1 ☐	2 ☐

i. Further observations:

20. In your experience, to what extent were the following methods used?

(Please circle the appropriate number)

	Not at all	Some (Less than 25%)	Often (25 - 50%)	Most of the time (Greater than 50% but not 100%)	Solely (100%)
a. Lectures	1	2	3	4	5
b. Discussion Groups (of attendees)	1	2	3	4	5
c. Panel Discussions	1	2	3	4	5
d. Role-playing or simulation	1	2	3	4	5
e. Other: (specify) _____ _____	1	2	3	4	5

21. Which aspect of the course did you find most helpful?

(check one)

1 ☐ Lectures

2 ☐ Discussion Groups

3 ☐ Handout Materials, Text

4 ☐ Simulation or Role-playing

5 ☐ Question and Answer Periods After Lecture

6 ☐ Informal Conversations with Other Participants

22. How would you assess the general quality of training—its strengths and weaknesses—provided by:

_____

(please fill in)

APPENDIX 5

Examples of Questionnaires Used in
the Terminal Evaluation of a Course or Workshop

(A) Form used to evaluate the Eleventh Annual Clinic on Library Applications of Data Processing, University of Illinois, May 1974.

Now that this Clinic is drawing to a close, we are certain that you have some reactions as to what aspects have been most valuable to you and what might have been different. As part of our effort to increase the effectiveness of future clinics, the planners would appreciate your completing this form. Your reactions will be confidential.

1. Please indicate your major purpose(s) for attending this Clinic.

   a. _____

   b. _____

   c. _____

   d. _____

2. What specific aspects of the Clinic experience contributed to your achieving one or any of the above personal objectives?

   _____

   _____

   _____

3. What specific aspects of the Clinic experience detracted from your achieving one or any of your previously stated objectives for attending the Clinic?

   _____

   _____

   _____

4. Would you give us your opinions in the following areas? (Check one in each area)

   a. Content of presentations
      Excellent . . . . . 1( )
      Satisfactory . . . . . 2( )
      Unsatisfactory . . . . . 3( )

   b. Number of presentations
      More needed . . . . . 1( )
      About right . . . . . 2( )
      Less needed . . . . . 3( )

c. Discussion and question period
   - More time needed . . . . . 1( )
   - About right . . . . . 2( )
   - Less time needed . . . . . 3( )

d. Size of group at this Clinic
   - Too many participants . . . . . 1( )
   - Size about right . . . . . 2( )
   - Too few participants . . . . . 3( )

e. Quality of presentations
   - Excellent . . . . . 1( )
   - Satisfactory . . . . . 2( )
   - Unsatisfactory . . . . . 3( )

f. Level of treatment
   - Above my head . . . . . 1( )
   - About right . . . . . 2( )
   - Too simple . . . . . 3( )

5. There are many parts of a Clinic experience that can either contribute to your satisfaction or detract from it. For each of the following, would you let us know how satisfied you have been. If any do not apply (e.g., you did not use hotel facilities), please leave blank.

a. Arrangements for coffee breaks
   - Really outstanding . . . . . 1( )
   - Very satisfactory . . . . . 2( )
   - Average . . . . . 3( )
   - Just acceptable . . . . . 4( )
   - Need improvement . . . . . 5( )

b. Hotel Rooms
   - Really outstanding . . . . . 1( )
   - Very satisfactory . . . . . 2( )
   - Average . . . . . 3( )
   - Just acceptable . . . . . 4( )
   - Need improvement . . . . . 5( )

c. Meeting Rooms
   - Really outstanding . . . . . 1( )
   - Very satisfactory . . . . . 2( )
   - Average . . . . . 3( )
   - Just acceptable . . . . . 4( )
   - Need improvement . . . . . 5( )

d. Organized social events (i.e., opening dinner and social hour)
   - Really outstanding . . . . . 1( )
   - Very satisfactory . . . . . 2( )
   - Average . . . . . 3( )
   - Just acceptable . . . . . 4( )
   - Need improvement . . . . . 5( )

6. In your opinion, how well did the total set of presentations cover the topic designated for this particular Clinic? (Check one)

Comprehensive Coverage					Inadequate Coverage
5	4	3	2	1	

7. In comparison with other continuing education activities which you have attended during the past few years, how would you rate this Clinic?

Excellent					Very Poor
5	4	3	2	1	

8. Prior to your attending, in what ways did you hear about this Clinic? (Check all that apply)

Received a brochure through the mail . . . . . 1( )
Saw news item in journals . . . . . 2( )
Through a friend or acquaintance . . . . . 3( )
Other_____ . . . . . 4( )
(Specify)

9. Please read all of the following statements. Then, check any of those that state how well you feel about the Clinic as a whole.

It was one of the most rewarding experiences I have ever had . . . . . 1( )
Exactly what I wanted . . . . . 2( )
It provided the kind of experience I can apply to my own situations . . . . . 3( )
It helped me personally . . . . . 4( )
It solved some problems for me . . . . . 5( )
I think it served its purpose . . . . . 6( )
It has some merits . . . . . 7( )
It was fair. . . . . 8( )
It was neither very good nor very poor . . . . . 9( )
I was mildly disappointed . . . . . .10( )
It was not exactly what I needed . . . . . .11( )
It was too general . . . . . .12( )
I am not taking any new ideas away . . . . . .13( )
It didn't hold my interest . . . . . .14( )
It was too superficial . . . . . .15( )
It was very poorly planned . . . . . .16( )
I didn't learn a thing . . . . . .17( )
It was a complete waste of time . . . . . .18( )

10. Would you be interested in attending another Clinic of this kind next year?

_____ Yes
_____ No

11. What topics do you recommend for next year's Clinic?

a. _____
b. _____
c. _____
d. _____

Thank you for your cooperation!

(B) End-of-course evaluation questionnaire used in the UNISIST Course on Information Retrieval and Information Retrieval Systems, Katowice, Poland, August 1974.

END-OF-COURSE EVALUATION

1. When you decided to attend this course you had certain objectives regarding what you hoped to gain from it. Circle the statement that best describes how far the course has met these objectives.

5	4	3	2	1
Completely	Almost completely	Partially	Incompletely	Very incompletely

2. In terms of your own professional goals, indicate how valuable this course is likely to be to you when you return to your own country.

5	4	3	2	1
Extremely valuable	Quite valuable	Of some value	Of little value	Of no value at all

3. Circle the one statement that best describes your assessment of the quality of the instruction as a whole.

5	4	3	2	1
Very good	Good	Average	Poor	Very poor

4. Would you please give us your evaluation on each of the following aspects? Check one box for each aspect.

(a) Discussion and question periods

- More time needed ☐
- About right ☐
- Less time needed ☐

(b) Student practical sessions

- More time needed ☐
- About right ☐
- Less time needed ☐

(c) Size of course

- Too many participants ☐
- Size about right ☐
- Too few participants ☐

(d) Level of treatment

Subject-matter too complex ☐

Treatment about right ☐

Treatment too simple ☐

5. There are many factors that may contribute to the satisfaction of participants in an educational experience of this type. For each of the following factors would you indicate how satisfied you have been?

(a) Arrangements for morning and afternoon breaks

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(b) Hotel accommodations

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(c) Lecture rooms

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

(d) Organized social events

Very good ☐

Satisfactory ☐

Unsatisfactory ☐

6. In your opinion how well did the complete sequence of lectures cover the subject of information retrieval?

Comprehensive coverage ☐

Satisfactory coverage ☐

Inadequate coverage ☐

7. Select the statement that best indicates the novelty of the material presented in this course.

   Most of the material was new to me ☐

   About half was new to me ☐

   Less than half was new to me ☐

   None of the material was new to me ☐

8. If another Unesco course of this type is conducted in the future would you recommend it to your colleagues?

   YES ☐

   NO ☐

9. If Unesco organizes a more advanced course in this subject area would you be interested in attending?

   YES ☐

   NO ☐

10. List below any suggestions you have as to how this course might have been improved.

APPENDIX 6

Follow-up Questionnaire on UNISIST International Training Course
on Information Retrieval and Information Retrieval Systems
Katowice, Poland, August 1974.

The course you attended in Katowice some months ago can be considered to have had a number of separate but closely related objectives:

1. To present a broad overview of the field of information retrieval with special reference to procedures used in modern information services.

2. To give participants a better understanding of what subject-matter needs to be covered in a curriculum in the field of information retrieval.

3. To illustrate in detail how a particular course in this subject might be conducted, including considerations of subject content, organization, teaching methods and teaching materials.

4. To stimulate participants to give more serious consideration to the problems of information retrieval in developing countries and to ways in which modern information handling systems might be introduced into these countries.

We would like your retrospective view of how well the Katowice course fulfilled these various objectives. Circle a number on each scale to indicate your assessment of the degree to which the course satisfied each of the four objectives (the higher numbers indicate greater satisfaction).

	Course completely satisfied this objective						Course contributed nothing to this objective
Objective 1:	7	6	5	4	3	2	1
Objective 2:	7	6	5	4	3	2	1
Objective 3:	7	6	5	4	3	2	1
Objective 4:	7	6	5	4	3	2	1

We would now like you to consider in what ways this particular UNISIST course has contributed to you professionally. Please indicate in what way, if any, you have been able to make use of the learning acquired during this course. Please check all appropriate boxes.

Have drawn upon the course in the planning of educational programs in my own country ☐

Have drawn upon the course in the development of a curriculum in the information sciences area ☐

Have used material gained from the course in my own teaching ☐

Was able to draw upon information gained from the course in the development of information services in my own country ☐

Have benefited from the course in other ways ☐
Please explain:

Have been unable to make any use of the material presented in this course ☐
Please explain why:

Please indicate on the scale below how valuable you feel this course has been to you professionally:

Great value						No value
7	6	5	4	3	2	1

If you have any further comments to make on the Katowice course please record these below.

Unesco is very grateful for your co-operation in completing this brief questionnaire. Please return it when completed to:

APPENDIX 7

Examples of Reports of Evaluations

(A) Report on the Evaluation of the Eleventh Annual Clinic on Library Applications of Data Processing, University of Illinois, May 1974.

Introduction

The eleventh annual Clinic on Library Applications of Data Processing was held by the Graduate School of Library Science, University of Illinois, from April 28 to May 1, 1974. The subject of this clinic, which was attended by approximately 140 librarians and others concerned with data processing applications, was minicomputers and their use in libraries. An attempt was made to evaluate the success of the clinic by the use of a questionnaire administered to all participants attending the final session of the clinic in May. The questionnaire used is reproduced in the appendix to this report. Questionnaires were received from 79 participants and the responses recorded are summarized in the remainder of this report.

Reasons for attending the clinic

The majority of the respondents (42/79) indicated that they attended the clinic in order to obtain a state-of-the-art review on the use of minicomputers in libraries. However, a substantial number (23/79) said that their main reason for attending was to meet other people working in the field, to find out what these people are doing, and to discover their opinions on various aspects of the topic of the clinic.

Major factors contributing to the attainment of objectives

The aspect of the clinic most frequently mentioned as contributing to the attainment of the personal objectives of participants was "conversations with those attending", mentioned by 25 of the respondents. As far as the formal aspects of the clinic are concerned, the group of papers from the University of Minnesota was the session most frequently mentioned (by 21 of the respondents). Mentioned six times each were the Curley presentation, the Beaumont presentation, the Davison presentation, and the tutorial by Corey and Divilbiss. "Social events" was mentioned five times by respondents and this could be grouped with the category "conversations with those attending" since, presumably, the main value of the social events lies in the opportunity provided for conversation with other participants.

Factors detracting from the attainment of objectives

The factor most frequently mentioned as a detraction was the quality and number of the audiovisual aids: fifteen respondents mentioned that these were inadequate or of poor quality. The remaining criticisms were conflicting. For example, a number of people said that there were too many "how to do it" talks and an approximately equal number said there were too few papers of this type and too many "promotional" presentations (this despite the fact that only two of the papers were given by representatives of commercial organizations).

Nine people suggested that the clinic would be better if the papers were available in advance. Five people suggested that the evening presentations should be eliminated to allow more time for informal contacts among participants.

Content of presentations

Excellent	21	28%
Satisfactory	42	56%
Unsatisfactory	0	
Variable (some satisfactory, some not)	12	16%
Nonrespondents	4	
TOTAL	79	100

Number of presentations

More needed	7	10%
About right	61	84%
Less needed	5	7%
Nonrespondents	6	
TOTAL	79	101

Discussion and question period

More time needed	16	22%
About right	57	77%
Less time needed	1	1%
Nonrespondents	5	
TOTAL	79	100

Quality of presentations

Excellent	15	20%
Satisfactory	44	59%
Unsatisfactory	4	5%
Quality varied	12	16%
Nonrespondents	4	
TOTAL	79	100

Level of treatment

Above my head	7	9%
About right	50	67%
Too simple	4	5%
Level varied	14	19%
Nonrespondents	4	
TOTAL	79	100

Coffee breaks

Really outstanding	7	9%
Very satisfactory	37	49%
Average	26	35%
Just acceptable	2	3%
Need improvement	3	4%
Nonrespondents	4	
TOTAL	79	100

Hotel rooms

Really outstanding	20	28%
Very satisfactory	36	50%
Average	14	19%
Just acceptable	2	3%
Need improvement	0	
Nonrespondents	7	
TOTAL	79	100

Meeting rooms

Really outstanding	10	13%
Very satisfactory	33	43%
Average	26	34%
Just acceptable	5	6%
Need improvement	3	4%
Nonrespondents	2	
TOTAL	79	100

Organized social events

Really outstanding	5	7%
Very satisfactory	33	46%
Average	23	32%
Just acceptable	9	12%
Need improvement	2	3%
Nonrespondents	7	
TOTAL	79	100

Extent of coverage of clinic topic

Comprehensive				Inadequate
8	45	13	10	0

Comparison with other continuing education activities attended in the past few years

Excellent				Very poor
17	37	14	3	0

How the attendee heard about the clinic

Received a brochure through the mail	41
Saw news item in journal	36
Through a friend or acquaintance	20
Other	15
Nonrespondents	2
	114*

*These are not mutually exclusive categories

General reaction to the clinic (Kropp-Verner Scale)

One of the most rewarding experiences I have ever had	4	5%
Exactly what I wanted	13	17%
Provided the kind of experience I can apply to my own situation	39	51%
It helped me personally	35	45%
It solved some problems for me	24	32%
I think it served its purpose	42	54%
It has some merits	12	16%
It was fair	2	3%
It was neither very good nor very poor	7	9%
I was mildly disappointed	14	18%
It was exactly what I needed	7	9%
I am not taking any new ideas away	1	1%
It didn't hold my interest	1	1%
It was too superficial	1	1%
It was very poorly planned	0	
I didn't learn a thing	0	
It was a complete waste of time	0	
Nonrespondents	2	
	206*	**

*The categories are not mutually exclusive

**These are percentages of the 77 respondents who checked each answer.

Would participant attend another clinic next year?

Yes	70	95%
No	4	5%
Nonrespondents	5	
TOTAL	79	100

Suggested topics for future clinics

Management, staffing, training	10
Cooperation and networking	10
Costs and economics	8
Systems for smaller libraries	7
Description and evaluation of operational systems	7
Patron use, reaction and training	6
Information retrieval	6
Serials control	5
New hardware	4
Data bases and data base services	4
Processing centers and services	3
Standards for automation	3
Changing role of the librarian	2

Conclusions and recommendations

On the whole it seems fair to say that this clinic was successful and that the great majority of those attending were well satisfied with the content and organization. There are many indications, however, that the informal aspects of a meeting of this type are at least as important as the formal aspects. A fairly substantial group of respondents indicated the importance of the social functions and the value of meeting with others working in the field. Others indicated a need to provide more time for discussions on an informal basis. In view of this reaction I suggest that we eliminate formal evening sessions at the next clinic except for the opening session on Sunday evening. The other two evenings will either be completely free or will be reserved for demonstrations or discussion periods among smaller groups of participants.

It also seems clear that we should try to exert more influence over participants in terms of their use of audiovisual materials. Although some were very good in this clinic (e.g., most of the BIBNET material), others were poor. In future I propose, when inviting further speakers, to stress the importance of good audiovisuals and to point out the size of type needed for adequate legibility in a group of the size we can expect.

A number of attendees expressed a desire for the papers to be printed and available in advance of the clinic or, at least, at the time the clinic is held. I am somewhat lukewarm about this idea. It is hard enough to get people to agree to present a paper and then to meet our publication deadline. It would, I feel, be much more of a problem to get them to prepare a formal paper in advance of the meeting. Moreover, it has been my experience that, in a large meeting, only a small proportion of those attending have read the papers in advance. Predistribution of papers works well with a small conference of invited participants (about 25 or 30 people) but is a less attractive proposition, I feel, for a group of over a hundred.

There is little evidence, from the results of this survey, that the size of the audience was considered too large. I would therefore suggest that we have no real need to restrict registration to a hundred, but should be able to go up to 125 or 150 without adverse effect on the clinic.

Finally, I believe that this formal evaluation has been useful and that we should continue to evaluate future clinics using approximately the same questionnaire. This will give us some standard of comparison from year to year.

APPENDIX 7

Examples of Reports
of Evaluations

(B)

****
* EVALUATION REPORT *
****

Third Annual ELECTRIC UTILITY WORKSHOP:
Water Treatment and Corrosion Control

March 14-16, 1983

Sponsored by:
State Water Survey Division of the
Illinois Department of Energy and Natural Resources
and
Division of Conferences and Institutes
Office of Continuing Education and Public Service
University of Illinois at Urbana-Champaign

by
Noreen Michael
Department of Educational Psychology
University of Illinois at Urbana-Champaign

for
Mrs. Mary Lewis, Program Assistant
Division of Conferences and Institutes
and
Program Committee
Electric Utility Workshop

May 1983

(Reproduced by permission of the
University of Illinois)

TABLE OF CONTENTS

PURPOSE	1
MAJOR EVALUATION QUESTIONS	1
BACKGROUND AND DESCRIPTION	1
SAMPLE OF PARTICIPANTS	2
INSTRUMENTS	2
DESCRIPTION OF FINDINGS	2
ANALYSIS AND SUMMARY	7
RECOMMENDATIONS	8

LIST OF APPENDICES

Appendix A	Workshop Schedule
Appendix B	Questionnaire
Appendix C	Interview Questions
Appendix D	Workshop Brochure

PURPOSE:

The purpose of the evaluation which was undertaken was to obtain information from participants at the Third Annual Electric Utility Workshop which will assist the management of the Office of Conferences and Institutes and the Program Committee of the Annual Electric Utility Workshop in planning future workshops to meet the needs of potential participants.

MAJOR EVALUATION QUESTIONS:

The main body of this evaluation report revolves around a two-page questionnaire. In addition, brief, informal interview questions were used during the two-day workshop. Due to the length of the questionnaire, specific questions will not be referred to at this point. However, a general idea of the issues which shaped the evaluation will be useful as an advance organizer. The questions asked in collecting evaluation data dealt primarily with:

1) Speaker preparedness and delivery
2) Written material quality and utility
3) Workshop content and level
4) Participant expectations and purpose(s) for attending workshop.

BRIEF BACKGROUND AND DESCRIPTION OF THE WORKSHOP:

The Third Annual Electric Utility Workshop is, as the title indicates, the third such workshop. As described in the 1983 brochure, the workshop is

> a comprehensive two day program designed to provide participants with a greater understanding of the latest techniques of controlling corrosion and improving the efficiency of operation by employing proper methods of water treatment, sampling, monitoring, and all means of instrumentation and how these techniques can be directly applied in the power plant.

This type of program is the only one of its kind in this region. It has steadily grown in numbers over the last three years. Last year there were 121 participants; this year there were 184, more than double the attendance at the first workshop. An added feature of this year's workshop was a social event which was planned as an "ice-breaker" and a time for workshop participants to meet each other. The event, a cocktail party, took place the night prior to the actual workshop sessions. A schedule of the two day workshop can be found in Appendix A.

SAMPLE OF PARTICIPANTS:

As mentioned previously, two methods of data collection were utilized -- a two-page questionnaire and informal interviews. In terms of sampling, provisions were made for all participants to complete the two page questionnaire. Prior to the commencement of the second day session, staff members from the Office of Conferences and Institutes placed evaluation forms at each seating position. On the other hand, because of the comprehensive nature of the workshop, it became difficult for the evaluator to select out specific persons for the informal interview component of the data gathering. During the coffee breaks most participants were engaged in conversation with others. A total of 30 participants were interviewed beginning at the three coffee breaks and at the break between the presentations and the Roundtables. Participants who sat in both smoking and no-smpking areas were interviewed; participants who sat either close to or far from the front of the room; as well as both male and female participants.

INSTRUMENTS:

The questionnaire used was designed from a question pool which had been generated in 1979 by a former staff member of the Office of Conferences and Institutes. The questionnaire was comprised of 26 structured questions and four open-ended questions. Structured questions all had scale values. A computerized form was used to avoid hand tabulation of structured questions. (See Appendix B)

The interview schedule consisted of four questions which related to the delivery and style of the speakers, the notebook, the expectations of the participants, and the Public Address System -- this question was included since the evaluator had problems hearing speakers. (See Appendix C)

D E S C R I P T I O N O F F I N D I N G S

In reporting the evaluation findings, data from the informal interviews will be interwoven to complement as well as supplement questionnaire data. A total of 123 participants responded to the questionnaire. This number represents 67% of the total number of participants -- 184. Frequencies will be reported when presenting the findings.

Questions 1 - 10 were answered accordingly: 1= POOR 2= INADEQUATE 3= ADEQUATE 4= VERY ADEQUATE 5= EXCELLENT

No.	QUESTION	Frequencies 1	2	3	4	5	NR*	X**
1.	The publicity for the workshop was	2	4	65	38	13	1	3.46
2.	The materials which I received in advance of the workshop were	4	23	54	33	7	2	3.13
3.	The accomodations were	2	12	38	53	16	2	3.57
4.	The opportunities for learning provided by the workshop were	1	4	41	56	21	0	3.75
5.	The opportunities I had for interaction with workshop faculty and participants were ...	3	8	42	49	21	0	3.63
6.	As a means of setting the stage for the workshop sessions, the cocktail party was ...	6	12	47	19	4	35	3.03
7.	The content of the workshop was ...	1	9	29	55	25	4	3.79
8.	The workshop facilities were ...	1	5	48	54	15	0	3.63
9.	The methods of presentation used during the workshop were	0	12	61	46	3	1	3.33
10.	The workshop speakers were ...	2	9	43	52	14	3	3.56

Questions 11 - 15 were answered accordingly: 1= STRONGLY DISAGREE 2= DISAGREE 3= UNSURE 4= AGREE 5= STRONGLY AGREE

No.	QUESTION	Frequencies 1	2	3	4	5	NR	X
11.	The presentations adequately represented the topics listed in the workshop brochure.	1	9	17	81	13	2	3.79
12.	I have gained new insights into my own situation from this workshop.	1	3	27	66	26	0	3.92
13.	I will be able to apply part of what I learned at this workshop to specific job situations.	0	3	21	76	23	0	3.97
14.	The workshop was worth my (company's) time, effort, and money.	1	6	8	70	37	1	4.11
15.	The speakers were thorough in their presentations.	0	7	30	75	9	2	3.71

*None responses ** Mean

Questions 16 - 19 asked participants to rate the workshop on various dimensions.

		Frequency	Percentage
16.	The level of the workshop was:		
	1= TOO ELEMENTARY FOR ME	14	11
	2= JUST RIGHT FOR ME	101	82
	3= TOO ADVANCED FOR ME	8	7
17.	The length of the workshop was:		
	1= TOO SHORT	26	21
	2= ABOUT RIGHT	50	41
	3= TOO LONG	45	37
	NR	2	2
18.	Would you recommend this workshop to others at your place of employment?		
	YES	114	93
	NO	8	7
	NR	1	1
19.	How did you receive your brochure for this workshop?		
	1= DIRECTLY IN THE MAIL	56	46
	2= FROM MY BOSS	30	24
	3= FROM A CO-WORKER	23	19
	4= REQUESTED ONE IN A PHONE CALL	4	3
	5= OTHER	2	2
	NR	8	7

Questions 20 - 26 were answered accordingly:

1 = MONDAY (PM) Social Hour as an ice-breaker activity.
2 = TUESDAY (AM), Analytical Chemistry Panel as a group.
3 = TUESDAY (PM), "Ion Exchange" -- Dr. Sallie Fisher.
4 = TUESDAY (PM), "Cooling Water Treatment" -- Paul Puckorius.
5 = TUESDAY (PM), Roundtable Discussions as a group.
6 = WEDNESDAY (AM), Instrumentation Panel as a group.
7 = WEDNESDAY (AM), Illinois Utility Projects Report Panel as a group.
8 = WEDNESDAY (PM), "Chemical Clearing" -- Gary Bradley

No.	QUESTION	Frequencies 1	2	3	4	5	6	7	8	N*	NR
20.	Which session will help you the most immediately upon return to your job?	1	20	36	27	12	10	2	5	0	10
21.	Which session did you feel was the most difficult to understand?	7	41	6	17	5	3	2	2	0	40
22.	Which session did you feel you benefited from the most?	0	17	42	21	15	13	2	3	0	10
23.	Which session did you enjoy the most?	3	6	51	24	14	8	3	2	0	12
24.	Which session, if any, could have been left out of the course?	2	20	3	7	2	1	10	1	77	0
25.	Which session, if any, could have been shorter?	2	34	4	7	0	7	8	3	58	0
26.	Which session, if any, could have been longer?	1	8	20	23	13	5	1	0	52	0

*None

OPEN-ENDED QUESTIONS:

For each open-ended question, a frequency distribution will be given of the participants comments classified by the most salient groupings. Following each distribution, representative comments will be presented.

27. What one thing did you enjoy most about the conference?

GROUPINGS	f	%	Participants' comments
Opportunity for interaction with peers	27	22.0	*Exchange of ideas. The interaction with people who think that chemistry in power plants is important. Talking with other people with common interests. The meeting of people in my field. General teaching talks by Fisher and Puckorius. S.P. Hall's presentation on Q.C. Analytic chemistry talks and handouts. Roundtable discussion. Roundtable -- feedwater-boiler water. The variety of speakers. Quality of most speakers. The level of experience of the speakers. I enjoyed the panel of speakers since the presentation was short and specific.*
A particular session	19	15.4	
Roundtables	18	14.6	
Quality/Variety of speakers	13	10.6	
Information received	12	9.6	
The materials	5	4.1	
Organization of workshop	2	1.7	
Other	3	2.4	
None responses	27	22.0	

28. What one thing did you least enjoy? How can we change this in the future? (The second part of this question is presented in a later section -- "Recommendations")

GROUPINGS	f	%	Participants' comments
Poor organization of notebook	15	12.2	*Too much was presented in too short a time frame. The fact that I only had enough time to attend 2 roundtable discussions. Some discussion periods were too short, causing the speakers to rush. Too much information packed into 2 days. Not enough breaks; too much sitting. No questions possible Tuesday AM. Pressed for time in almost all presentations. The long day Tuesday. The Analytic chemistry group was too detailed. Analytic chemistry panel seemed too elementary. Need information about newest systems for water and operations analyses. The brevity given to the analytic session. The analytic chemistry panel was too much of a sales pitch for particular instruments. The analytic session; this was to be a Utility Workshop about Plant problems. The book was ill prepared. The unorganized book. A lot of printed material is useless if it cannot be accessed. The various topics are in helter-skelter order with no table of contents or tabs! Those people that read their papers. Speakers reading notes we already have. Presenters who read their papers rather than made a comprehensive audio/visual presentation. Reading of papers*
Crammed presentation schedule; little time for question & answers..	13	10.6	
Analytic Chemistry session	10	8.1	
Reading of papers.....	8	6.5	
Mediocre quality of speakers	6	4.9	
Lack of adherence to schedule	7	5.7	
Poor Public Address System	6	4.9	
Redundancy of information	5	4.1	
Insufficient breaks ...	5	4.1	
First day session too long	5	4.1	
Information overload	4	3.3	
Cocktail party	3	2.4	
Miscellaneous	17	13.8	
None responses	27	22.0	

29. Why did you decide to attend this workshop?

GROUPINGS	f	%
1. To enrich /upgrade my knowledge of the content area	32	26.0
2. To obtain information to utilize in my work situation	22	17.8
3. Required by boss	7	5.7
4. Recommended by supervisor/coworker/ former participant	8	6.5
5. Brochure information	14	11.3
6. Beneficial nature of previous utility workshop(s)	6	4.9

The information obtained through the informal interviews support many of the points made by participants in questions 28 and 29 on the questionnaire. A majority of the interviewees found the notebook disorganized, though valuable.

Because of the nature of question 29, no direct quotations will be presented. The groupings adequately reflect the content of most comments. About 7 persons, 5.7% made comments other than those reflected in the table on the previous page. Twenty-three participants, 18.7% did not respond to this question.

30. How meaningful were the Roundtable sessions to you?

GROUPINGS	f	%	Participants' comments
Extremely meaningful/ Beneficial	17	13.8	*Very, very important. More helpful than any other session. The most meaningful. Fantastic; more time should be devoted to this. Best part of the conference. Very meaningful. I would like to have attended all of them. Very meaningful and should have been a day long. Good - Wish time would have allowed a real second session. I learned a lot from it. Could be better - how about separating some sessions to get groups of participants with like backgrounds. Not very; too hurried, too crowded; needs much more emphasis in conference.*
Very meaningful/ Good	29	23.6	
Meaningful/ Good	21	17.1	
Fairly meaningful/ Good	9	7.3	
Inadequate	13	10.6	
Too tired to attend .	6	4.9	
None responses	28	22.8	

ANALYSIS AND SUMMARY:

The data presented indicates that there was a mixed group in attendance at the workshop. For example, some participants enjoyed the panel on Tuesday morning; others thought it was repetitious. Some participants enjoyed and wanted more time for the Analytic Chemistry session; others thought it irrelevant and unnecessary. In addition, in the informal interviews conducted, the same type of response was received. Some participants thought that the Tuesday morning session was long and tedious and not very significant; others said that the information discussed in the session was exactly what they came to the workshop to receive. These examples point to a need for the Planning Committee of the Program Committee, in conjunction with the Division of Conferences and Institutes to consider modifying future workshops in an effort to meet the needs of most, if not all, future participants. Modification of the workshop may also help in attracting former participants.

One of the reasons for group diversity seems to stem from the level of on the job experience of the participants. Some participants indicated during the interview sessions that they were new at their jobs and much of what was said was over their heads. On the other hand, one participant mentioned that he had had a similar problem when he attended the workshop last year, but because of the exposure which he received on the job in the last year, he was able to grasp much more at this year's workshop, which was a more meaningful experience than was last year's.

Another reason for group diversity seemed to stem from the job responsibilities of participants. Again, during the interviews, several participants indicated that they were not interested in certain sessions, since those sessions were not relevant or specific to their job situation. This could be an explanation for the ratings that various sessions received, as depicted in the data presentation on the first half of page 5. These two factors -- on the job experience and job responsibilities -- also surfaced in various open-ended comments. Thus, the Program Committee and the Division of Conferences and Institutes need to carefully weigh these factors along with the findings reported in the previous pages when planning future workshops. In addition, the following recommendations, made by participants, should be considered.

RECOMMENDATIONS:

(Note: The following recommendations reflect part two of question 28, as well as comments from informal interviews.)

1. *Have speakers give overview of report rather than read report.*
2. *Organize the book better with index and indexing tabs.*
3. *Have roundtables both days.*
4. *The workshop should be three days.*
5. *Instruct speakers not to read papers.*
6. *Have Sallie Fisher one whole day.*
7. *More topics -- less repetition of topics.*
8. *Provide highliters in lieu of pens.*
9. *Have more frequent breaks.*
10. *Have cocktail party after first day of workshop rather than before.*
11. *Use larger name tags.*